

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 April 2003 (10.04.2003)

PCT

(10) International Publication Number
WO 03/029271 A2

(51) International Patent Classification⁷: **C07K**

(21) International Application Number: PCT/US02/30474

(22) International Filing Date:
24 September 2002 (24.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/324,631 24 September 2001 (24.09.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US Not furnished (CIP)
Filed on Not furnished

(71) Applicant (for all designated States except US): **HYSEQ, INC.** [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94085 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **TANG, Tom, Y.** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **ZHANG, Jie** [CN/US]; 4930 Poplar Terrace, Campbell, CA 95008 (US). **REN, Feiyan** [US/US]; 7703 Oak Meadow Court, Cupertino, CA 95014 (US). **XUE, Aidong, J.** [US/US]; 1621 S. Mary Avenue, Sunnyvale, CA 94087 (US). **ZHAO, Qing, A.** [CN/US]; 1556 Kooser Rd, San Jose, CA 95118 (US). **WANG, Jian-Rui** [CN/US]; 744 Stendhal Lane, Cupertino, CA 95014 (US). **WEHRMAN, Tom** [US/US]; CCSR Mol Pharm 3210, 269 W. Campus Dr, Stanford, CA 94305 (US). **ZHOU, Ping** [US/US]; 7595 Newcastle Drive, Cupertino, CA 95014 (US). **GHOSH, Malabika** [IN/US]; 620 Iris Avenue, Apt. 133, Sunnyvale, CA 94086 (US). **WANG, Dunrui** [CN/US]; 12252 Pepper Tree Lane, Poway, CA 92064 (US). **MA, Yunqing**

[CN/US]; 280 W. California Ave #206, Sunnyvale, CA 94086 (US). **ASUNDI, Vinod** [US/US]; 709 Foster City Boulevard, Foster City, CA 94404 (US). **WANG, Zhiwei** [CN/US]; 836 Alturas Ave #B36, Sunnyvale, CA 94085 (US). **WENG, Gezhi** [US/US]; 415 Moraga Avenue, Piedmont, CA 94611 (US). **HALEY-VICENTE, Dana** [US/US]; 1305 Eagle Glen, Escondido, CA 92029 (US). **DRMANAC, Rodoje, T.** [US/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agents: **RIN-LAURES, Li-Hsien** et al.; Hyseq, Inc., 670 Almanor Avenue, Sunnyvale, CA 94085 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by the nucleic acids and uses thereof.



WO 03/029271 A2

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of U.S. Provisional Application Serial No. 60/324,631 filed September 24, 2001 entitled "Novel Nucleic Acids and Polypeptides", Attorney Docket No. 810, which is a continuation-in-part application of PCT Application Serial No. PCT/US00/35017 filed December 22, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784CIP3A/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/552,317 filed April 25, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/488,725 filed January 21, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 784; PCT Application Serial No. PCT/US01/02623 filed January 25, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 785CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/491,404 filed January 25, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 785; PCT Application Serial No. PCT/US01/03800 filed February 5, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/560,875 filed April 27, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/496,914 filed February 03, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 787; PCT Application Serial No. PCT/US01/04927 filed February 26, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/577,409 filed May 18, 2000 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 788CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/515,126 filed February 28, 2000 entitled "Novel Contigs Obtained from Various Libraries"; Attorney Docket No. 788; PCT Application Serial No. PCT/US01/04941 filed March 5, 2001 entitled "Novel Contigs Obtained from Various Libraries", Attorney Docket No. 789CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/574,454 filed May 19, 2000 entitled "Novel Contigs Obtained from Various

Libraries”, Attorney Docket No. 789CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/519,705 filed March 07, 2000 entitled “Novel Contigs Obtained from Various Libraries”, Attorney Docket No. 789; PCT Application Serial No. PCT/US01/08631 filed March 30, 2001 entitled “Novel Contigs Obtained from Various Libraries”, Attorney Docket No. 790CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/649,167 filed August 23, 2000 entitled “Novel Contigs Obtained from Various Libraries”, Attorney Docket No. 790CIP, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/540,217 filed March 31, 2000 entitled “Novel Contigs Obtained from Various Libraries”, Attorney Docket No. 790; PCT Application Serial No. PCT/US01/08656 filed April 18, 2001 entitled “Novel Contigs Obtained from Various Libraries”, Attorney Docket No. 791CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/770,160 filed January 26, 2001 entitled “Novel Contigs Obtained from Various Libraries”, Attorney Docket No. 791CIP, which is in turn a continuation-in-part application of U.S. Application Serial No. 09/552,929 filed April 18, 2000 entitled “Novel Contigs Obtained from Various Libraries”, Attorney Docket No. 791; and PCT Application Serial No. PCT/US01/14827 filed May 16, 2001 entitled “Novel Contigs Obtained from Various Libraries”, Attorney Docket No. 792CIP3/PCT, which in turn is a continuation-in-part application of U.S. Application Serial No. 09/577,408 filed May 18, 2000 entitled “Novel Contigs Obtained from Various Libraries”, Attorney Docket No. 792; all of which are incorporated herein by reference in their entirety.

2. BACKGROUND OF THE INVENTION

2.1 TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

2.2 BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, circulating soluble factors, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression

cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence
5 cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader
10 sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types
15 of data and products dependent on DNA and amino acid sequences.

3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules,
20 cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including
25 expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by
30 hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1-971, or 1943-2709 and are provided in

the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases or unknown. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences
5 that hybridize to the complement of SEQ ID NO: 1-971, or 1943-2709 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1-971, or 1943-2709. A polynucleotide comprising a nucleotide sequence having at
10 least 90% identity to an identifying sequence of SEQ ID NO: 1-971, or 1943-2709 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-971, or 1943-2709. The
15 sequence information can be a segment of any one of SEQ ID NO: 1-971, or 1943-2709 that uniquely identifies or represents the sequence information of SEQ ID NO: 1-971, or 1943-2709.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information are
20 provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences;
25 and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing full-length genes, use for chromosome and gene mapping, use in the
30 recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-971, or 1943-2709 or novel segments or parts of the nucleic acids of the invention are used as primers in

expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-971, or 1943-2709 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed
5 sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1-971, or 1943-2709; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1-971, or 1943-2709; and a polynucleotide comprising any of the nucleotide
10 sequences of the mature protein coding sequences of SEQ ID NO: 1-971, or 1943-2709. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1-971, or 1943-2709; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in SEQ ID NO: 1-971, or 1943-2709;
15 (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homologue (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in SEQ ID NO: 972-1942, or 2710-3476, or Tables 3, 4A, 4B, or 5.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in the Sequence Listing; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-971, or 1943-2709; or (b) polynucleotides that
25 hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially
30 chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

5 The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in
10 which the protein produced by such processes is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in
15 generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, e.g., *in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as
20 expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a
25 polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical
30 condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the
5 polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a
10 complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under
15 conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs,
20 and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the
25 identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (e.g., bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell
30 for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound that binds to a polypeptide of the invention is identified.

The methods of the invention also provide methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of the target gene products. Compounds and other substances can affect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Tables 2A and 2B); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Tables 4A and 4B). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

4. DETAILED DESCRIPTION OF THE INVENTION

4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms “a”, “an” and “the” include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms “biologically active” or “biological activity” refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise “immunologically active” or “immunological activity” refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms “complementary” or “complementarity” refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded

molecules may be "partial" such that only certain portion(s) of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

5 The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell
10 lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are
15 able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF.

20 EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or
25 "oligonucleotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and
30 N is A, C, G, or T (U) or unknown. It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of

oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NO: 1-971, or 1943-2709.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-971, or 1943-2709. The sequence information can be a segment of any one of SEQ ID NO: 1-971, or 1943-2709 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-971, or 1943-2709, or those segments identified in Tables 3, 4A, 4B, or 5. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-

mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully
5 matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

10 Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1/4^{25}$) times the increased probability for mismatch at each nucleotide position (3×25). The probability that an eighteen mer with a single mismatch can be detected in an array for
15 expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

20 The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

25 The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally
30 occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids,

more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full-length protein which may include any leader sequence or any processing sequence.

The term "mature protein coding sequence" means a sequence which encodes a peptide or protein without a signal or leader sequence. The "mature protein portion" means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant" (or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, e.g., recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular

prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

5 Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino
10 acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids,
15 more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or
20 non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate
25 polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological
30 macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present

(but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (e.g., microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the

regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (*e.g.* Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2): 134 -143) and factors released from damaged cells (*e.g.* Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (*i.e.*, hybridization to filter-bound DNA in 0.5 M NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (*i.e.*, washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligonucleotides), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" or "substantially similar" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of

which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more than 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, the nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, more preferably at least about 80% sequence identity, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least about 95% sequence identity, more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a new stop codon) should be disregarded. Sequence identity may be determined, *e.g.*, using the Jotun Hein method (Hein, J. (1990) *Methods Enzymol.* 183:626-645). Identity between sequences can also be determined by other methods known in the art, *e.g.* by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

4.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1-971, or 1943-2709; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO: 1-971, or 1943-2709; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1-971, or 1943-2709. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO: 1-971, or 1943-2709; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homologue of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 972-1942, or 2710-3476 (for example, as set forth in Tables 3, 4A, 4B, or 5). Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like

polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include
5 receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include entire coding region of the cDNA or may represent a portion of the coding region of the cDNA.

10 The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials.
15 Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1-971, or 1943-2709 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1-971, or 1943-2709 or a portion thereof as a probe. Alternatively, the polynucleotides of
20 SEQ ID NO: 1-971, or 1943-2709 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence
25 information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, e.g., at least about 65%, at least about
30 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, more typically at least about 85%, 86%, 87%, 88%, 89%, more typically at least about 90%, 91%, 92%, 93%, 94%, and even more typically at least about 95%, 96%, 97%, 98%, 99% sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1-971, or 1943-2709, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to) any one of the polynucleotides of the invention are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1-971, or 1943-2709, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NO: 1-971, or 1943-2709 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology results for the nucleic acids of the present invention, including SEQ ID NO: 1-971, or 1943-2709 can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST (Basic Local Alignment Search Tool) program is used to search for local sequence alignments (Altschul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using FASTXY algorithm may be performed.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification

results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

5 A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a
10 functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

 Polynucleotides encoding preferred polypeptide truncations of the invention could be
15 used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

 The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such
20 polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

 In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-971, or 1943-2709, or
25 functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

 A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et
30 al. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a

polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-971, or 1943-2709 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-971, or 1943-2709 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example: Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene), pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two

appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells

are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., Nat. Biotech 17, 870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intra-muscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

4.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1-971, or 1943-2709, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO: 1-971, or 1943-2709 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO: 1-971, or 1943-2709 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO: 1-971, or 1943-2709, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

4.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of an mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a

DNA disclosed herein (*i.e.*, SEQ ID NO: 1-971, or 1943-2709). For example, a derivative of Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, mRNA of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

4.5 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids

of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under

the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell* 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by

phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, and regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the

host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 972-1942, or 2710-3476 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NO: 1-971, or 1943-2709 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NO: 1-971, or 1943-2709 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 972-1942, or 2710-3476 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 972-1942, or 2710-3476 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, typically at least about 95%, 96%, 97%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 972-1942, or 2710-3476.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as

described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites. Fragments are also identified in Tables 3, 4A, 4B, or 5.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The predicted signal sequence is set forth in Table 6. The mature form of such protein may be obtained and confirmed by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell and sequencing of the cleaved product. One of skill in the art will recognize that the actual cleavage site may be different than that predicted in Table 6. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed (See, e.g., Sakal et al., Prep. Biochem. Biotechnol. (2000), 30(2), pp. 107-23, incorporated herein by reference).

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological

properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified
5 proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell,
10 through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising
15 growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is
20 cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange
25 chromatography, and immuno-affinity chromatography. See, e.g., Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in *Molecular Cloning: A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include
30

fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 972-1942, or 2710-3476.

The protein of the invention may also be expressed as a product of transgenic animals, e.g., as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, e.g., U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino

acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant

methyle or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

4.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), Pfam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobicity prediction

algorithm (J. Mol Biol, 157, pp. 105-31 (1982), the GeneAtlas software (Molecular Simulations Inc. (MSI), San Diego, CA) (Sanchez and Sali (1998) Proc. Natl. Acad. Sci., 95, 13597-13602; Kitson DH et al, (2000) "Remote homology detection using structural modeling – an evaluation" Submitted; Fischer and Eisenberg (1996) Protein Sci. 5, 947-955), Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark) incorporated herein by reference). Polypeptide sequences were examined by a proprietary algorithm, SeqLoc that separates the proteins into three sets of locales: intracellular, membrane, or secreted. This prediction is based upon three characteristics of each polypeptide, including percentage of cysteine residues, Kyte-Doolittle scores for the first 20 amino acids of each protein, and Kyte-Doolittle scores to calculate the longest hydrophobic stretch of the said protein. Values of predicted proteins are compared against the values from a set of 592 proteins of known cellular localization from the Swissprot database (<http://www.expasy.ch/sprot>). Predictions are based upon the maximum likelihood estimation.

Presence of transmembrane region(s) was detected using the TMPred program (http://www.ch.embnet.org/software/TMPRED_form.html).

The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

4.7 CHIMERIC AND FUSION PROTEINS

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide according to the invention and the other polypeptide are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus, or to the middle.

For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein.

In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies in a subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

4.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to
5 appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For
10 additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may
15 also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense
20 therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present
25 invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of
30 the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be

modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences.

5 See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with
10 the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control
15 of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment
20 regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion
25 properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, e.g., inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple
30 deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are

deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappell; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No. 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by
5 supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to
10 express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or
15 inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals,
20 can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No
25 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or
30 even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

4.10 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment.

Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

4.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA

sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

4.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid

preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

5 **4.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY**

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations.

10 A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell
15 proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those
20 described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology
25 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan
30 eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin- γ , Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

4.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotent or pluripotent state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors.

The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotent/pluripotent stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotent/pluripotent mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies

would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., *Differentiation*, 48: 173-182, (1991); Klug et al., *J. Clin. Invest.*, 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering* eds. Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. *Proc. Natl. Acad. Sci, U.S.A.*, 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the

invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

4.10.5 HEMATOPOIESIS REGULATING ACTIVITY

5 A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, 10 thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in 15 supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and 20 therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or 25 heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

30 Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

- Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994;
- 5 Hirayama et al., *Proc. Natl. Acad. Sci. USA* 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., *Experimental Hematology* 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In *Culture of Hematopoietic Cells*. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc.,
- 10 New York, N.Y. 1994.
- 15

4.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and

20 tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have

25 prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming

30 cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast

activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from

chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

4.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and

disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, *Leishmania* spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also be useful in the treatment of allergic reactions and conditions (e.g., anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., *Toxicology* 125: 59-66, 1998), skin prick test (Hoffmann et al., *Allergy* 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., *Arch. Toxicol.* 73: 501-9), and murine local lymph node assay (Kimber et al., *J. Toxicol. Environ. Health* 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of

an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing
5 non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without
10 limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition
15 as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may
20 avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in
25 humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed.,
30 Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self-tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro

antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

4.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in:

Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

4.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to

tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell

5 population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

10 Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E.
15 Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

20

4.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostasis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders
25 (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

30 Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis

Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

4.10.11 CANCER DIAGNOSIS AND THERAPY

5 Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing
10 malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation,
15 inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies
20 including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal
25 neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central
30 nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Kaposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention

(including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987)

Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis
5 assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

10

4.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such
15 receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and
20 humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured
25 by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions
30 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

5 Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182
10 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

15 4.10.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One
20 method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or
25 fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries
30 comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product
5 libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science* 282:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide
10 and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol.* 9(3):205-23 (1998); Hruby
15 et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then
20 tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The
25 toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

4.10.14 ASSAY FOR RECEPTOR ACTIVITY

30 The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening

assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of
5 compounds, and in particular small molecules, that modulate (*i.e.*, increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does
10 not. The responses of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic
15 chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the
20 extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications *i.e.* phosphorylation. Other methods known to those in the art can also be used to identify
25 signaling molecules involved in receptor activity.

4.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in
30 the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an

inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflammation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic myelogenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

4.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

4.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include

but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or

5 compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or
10 injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration
15 associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency,
20 Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

25 (vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various
30 etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival

or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

- (i) increased survival time of neurons in culture;
- (ii) increased sprouting of neurons in culture or *in vivo*;
- 5 (iii) increased production of a neuron-associated molecule in culture or *in vivo*,
e.g., choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or
- (iv) decreased symptoms of neuron dysfunction *in vivo*.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method
10 set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor
15 neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor
20 neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory
25 Neuropathy (Charcot-Marie-Tooth Disease).

4.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing,
30 infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution,

change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

4.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that

hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

4.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et al., 1983, *Science*, 219:56, or by B. Waksman et al., 1963, *Int. Arch. Allergy Appl. Immunol.*, 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed *Mycobacterium tuberculosis* in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed *Mycobacterium tuberculosis* in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of *Mycobacterium* CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

4.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

5

4.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01 µg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1 µg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

25 4.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other

30

materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF- α and TGF- β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound

sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

4.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in

fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

4.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical

composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from
5 about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally
10 acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride
15 Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or
20 physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such
25 carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable
30 excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired,

disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such

as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the
5 preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or
10 other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion
15 exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate
20 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics.
25 Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, *e.g.* polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds
30 may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable

matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides,

diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

5 The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient.

10 Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to

15 practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition

20 topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically

25 useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active

30 ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above-mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet

derived growth factor (PDGF), transforming growth factors (TGF- α and TGF- β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either *in vivo* or *ex vivo* into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

4.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be

estimated initially from appropriate *in vitro* assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as
5 determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic
10 efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50} . Compounds which exhibit high therapeutic
15 indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of
20 administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in*
25 *vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of
30 the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 µg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 µg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

4.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

4.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen-binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, F_{ab'} and F_{(ab')₂} fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for

polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 972-1942, or 2710-3476, or Tables 3, 4A, 4B, or 5, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a surface region of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

The term "specific for" indicates that the variable regions of the antibodies of the invention recognize and bind polypeptides of the invention exclusively (*i.e.*, able to distinguish the polypeptide of the invention from other similar polypeptides despite sequence identity, homology, or similarity found in the family of polypeptides), but may also interact with other proteins (for example, *S. aureus* protein A or other antibodies in ELISA techniques) through interactions with sequences outside the variable region of the antibodies, and in particular, in the constant region of the molecule. Screening assays to determine

binding specificity of an antibody of the invention are well known and routinely practiced in the art. For a comprehensive discussion of such assays, see Harlow et al. (Eds), Antibodies A Laboratory Manual; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Chapter 6. Antibodies that recognize and bind fragments of the polypeptides of the invention are also contemplated, provided that the antibodies are first and foremost specific for, as defined above, full-length polypeptides of the invention. As with antibodies that are specific for full length polypeptides of the invention, antibodies of the invention that recognize fragments are those which can distinguish polypeptides from the same family of polypeptides despite inherent sequence identity, homology, or similarity found in the family of proteins.

Antibodies of the invention are useful for, for example, therapeutic purposes (by modulating activity of a polypeptide of the invention), diagnostic purposes to detect or quantitate a polypeptide of the invention, as well as purification of a polypeptide of the invention. Kits comprising an antibody of the invention for any of the purposes described herein are also comprehended. In general, a kit of the invention also includes a control antigen for which the antibody is immunospecific. The invention further provides a hybridoma that produces an antibody according to the invention. Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Monoclonal antibodies binding to the protein of the invention may be useful diagnostic agents for the immunodetection of the protein. Neutralizing monoclonal antibodies binding to the protein may also be useful therapeutics for both conditions associated with the protein and also in the treatment of some forms of cancer where abnormal expression of the protein is involved. In the case of cancerous cells or leukemic cells, neutralizing monoclonal antibodies against the protein may be useful in detecting and preventing the metastatic spread of the cancerous cells, which may be mediated by the protein.

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics. The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and Sepharose®, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known

in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

4.13.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface-active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific

antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8
5 (April 17, 2000), pp. 25-28).

4.13.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular
10 species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen-binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

15 Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256, 495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be
20 immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell
25 line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that
30 preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas

typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107, 220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as

a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA
5 also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted
10 for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

4.13.3 HUMANIZED ANTIBODIES

15 The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab',
20 F(ab')₂ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321, 522-525 (1986); Riechmann et al., Nature, 332, 323-327 (1988); Verhoeyen et al., Science, 239, 1534-1536 (1988)), by substituting
25 rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539). In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise
30 substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion

of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, *Curr. Op. Struct. Biol.*, 2, 593-596 (1992)).

5 4.13.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed “human antibodies”, or “fully human antibodies” herein. Human monoclonal antibodies can be prepared by the trioma technique; the human
10 B-cell hybridoma technique (see Kozbor, et al., 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. *Proc Natl Acad Sci USA* 80,
15 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227, 381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by
20 introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806;
25 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (*Bio/Technology* 10, 779-783 (1992)); Lonberg et al. (*Nature* 368, 856-859 (1994)); Morrison (*Nature* 368, 812-13 (1994)); Fishwild et al, (*Nature Biotechnology* 14, 845-51 (1996)); Neuberger (*Nature Biotechnology* 14, 826 (1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13, 65-93 (1995)).

30 Human antibodies may additionally be produced using transgenic nonhuman animals that are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains

in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then
5 obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the XenomouseTM as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after
10 immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for
15 example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent
20 rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light
30 chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that

binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

4.13.5 FAB FRAGMENTS AND SINGLE CHAIN ANTIBODIES

5 According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246, 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or
10 derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab)2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab)2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing
15 agent and (iv) F_v fragments.

4.13.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of
20 the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two
25 immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305, 537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished
30 by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10, 3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion

preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121, 210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full-length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., *Science* 229, 81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab' -TNB derivatives is then reconverted to the Fab' -thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab' -TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175, 217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical

coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5), 1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90, 6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152, 5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., *J. Immunol.* 147, 60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG ($Fc\gamma R$), such as $Fc\gamma RI$ (CD64), $Fc\gamma RII$ (CD32) and $Fc\gamma RIII$ (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

4.13.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

4.13.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176, 1191-1195 (1992) and Shopes, J. Immunol., 148, 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53, 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3, 219-230 (1989).

4.13.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used

include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, *Aleurites fordii* proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

4.14 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the

presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NO: 1-971, or 1943-2709 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NO: 1-971, or 1943-2709 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein-encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the

present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means
5 having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present
10 invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target
15 sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available
20 algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The
25 most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally
30 selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include,

but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

4.15 TRIPLE HELIX FORMATION

5 In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are preferably 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple
10 helix-see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 15241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization
15 blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

20 4.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

25 In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization
30 conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

5 In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary.

10 Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay.

One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science

15 Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or

20 membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is

25 compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies
30 of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

4.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NO: 1-971, or 1943-2709, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and
5 detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting
10 the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression
15 of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to
20 activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in
25 the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

30 For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed"

when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al.,
5 Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or
10 EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple
15 helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix -
20 see Lee et al., Nucl. Acids Res. 6, 3073 (1979); Cooney et al., Science 241, 456 (1988); and Dervan et al., Science 251, 1360 (1991)) or to the mRNA itself (antisense-Okano, J. Neurochem. 56, 560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks
25 translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention
30 can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

4.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NO: 1-971, or 1943-2709. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from any of the nucleotide sequences SEQ ID NO: 1-971, or 1943-2709 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well-known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal

map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

5 4.20 **PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES**

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6), 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 15 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8), 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be 20 purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the 25 microwell surface termed CovaLink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridgeheads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal. Biochem. 198(1) 138-42).

30 The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen *et al.*, (1991). In this technology, a phosphoramidate bond is employed (Chu *et al.*, (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins

the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ μ l) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-MeIm₇), is then added to a final concentration of 10 mM 1-MeIm₇. A ss DNA solution is then dispensed into CovaLink NH strips (75 μ l/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-MeIm₇, is made fresh and 25 μ l added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995), 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res., 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal. Biochem. 169(1), 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) Proc. Nat'l. Acad. Sci., USA 91(11), 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

4.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24), 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *Cvi*JI, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *Cvi*JI normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*Cvi*JI**), yield a quasi-random distribution of DNA fragments from the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *Cvi*JI** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *Cvi*JI** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 µg instead of 2-5 µg); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed).

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

4.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm², depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient.

Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell
5 plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and
10 variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed,
15 numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated
20 by reference in their entirety.

5.0 EXAMPLES

5.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from
25 various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature
30 sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences.

5.2 EXAMPLE 2

Assemblage of Novel Nucleic Acids

The contigs or nucleic acids of the present invention, designated as SEQ ID NO: 1943-2709 were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST, gb pri, and UniGene, and exons from public domain genomic sequences predicated by GenScan) that belong to this assemblage. The algorithm terminated when there were no additional sequences from the above databases that would extend the assemblage. Further, inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

The novel predicted polypeptides (including proteins), SEQ ID NO: 2710-3476, encoded by the novel polynucleotides (SEQ ID NO: 1943-2709) of the present invention, and their corresponding translation start and stop nucleotide locations to each of SEQ ID NO: 1943-2709 were obtained using one of two methods. Polypeptides were obtained by using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which selects a polypeptide based on a comparison of the translated novel polynucleotide to known polynucleotides (W.R. Pearson, Methods in Enzymology, 183:63-98 (1990), herein incorporated by reference). Alternatively, polypeptides were obtained by using a software program called GenScan for human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a probabilistic model of gene structure/compositional properties (C. Burge and S. Karlin, J. Mol. Biol., 268:78-94 (1997), incorporated herein by reference). Method C refers to a polypeptide obtained by using a Hyseq proprietary software program that translates the novel polynucleotide and its complementary strand into six possible amino acid sequences (forward and reverse frames) and chooses the polypeptide with the longest open reading frame.

5.3 EXAMPLE 3

Novel Nucleic Acids

The novel nucleic acids of the present invention were assembled from sequences that were obtained from a cDNA library by methods described in Example 1 above, and in some cases sequences obtained from one or more public databases. The nucleic acids were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (Hyseq's database containing EST sequences, dbEST, gb pri, and UniGene) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full-length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequences were checked using FASTY and/or BLAST against Genbank (i.e., dbEST, gb pri, UniGene, and Genpept) and the Geneseq (Derwent). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Hyseq, Inc.). The full-length nucleotide and amino acid sequences, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NO: 1-1942.

Table 1 shows the various tissue sources of SEQ ID NO: 1-971.

The results showing homologues for SEQ ID NO: 972-1942 from Genpept 124 are shown in Table 2A. The results showing homologues for SEQ ID NO: 972-1942 from Genpept 129 are shown in Table 2B.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6, 219-235 (1999), <http://motif.stanford.edu/ematrix-search/> herein incorporated by reference), all the polypeptide sequences were examined to determine whether they had identifiable signature regions. Scoring matrices of the eMatrix software package are derived from the BLOCKS, PRINTS, PFAM, PRODOM, and DOMO databases. Table 3 shows the accession number of the homologous eMatrix signature found in the indicated polypeptide sequence, its description, and the results obtained which include accession number subtype; raw score; p-value; and the position of signature in amino acid sequence.

Using the Pfam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4A shows the name of the Pfam model found, the description, the e-value and the Pfam score for the identified model within the sequence as described in United States priority application serial number 60/324,631, filed September 24, 2001, herein incorporated by reference in its entirety. Table 4B shows the name of the Pfam model found, the description, the e-value and the Pfam score for the identified model within the sequence using Pfam version 7.2. Further description of the Pfam models can be found at <http://pfam.wustl.edu/>.

The GeneAtlas™ software package (Molecular Simulations Inc. (MSI), San Diego, CA) was used to predict the three-dimensional structure models for the polypeptides encoded by SEQ ID NO: 1-971 (i.e. SEQ ID NO: 972-1942). Models were generated by (1) PSI-BLAST which is a multiple alignment sequence profile-based searching developed by Altschul et al, (Nucl. Acids. Res. 25, 3389-3408 (1997)), (2) High Throughput Modeling (HTM) (Molecular Simulations Inc. (MSI) San Diego, CA,) which is an automated sequence and structure searching procedure (<http://www.msi.com/>), and (3) SeqFold™ which is a fold recognition method described by Fischer and Eisenberg (J. Mol. Biol. 209, 779-791 (1998)). This analysis was carried out, in part, by comparing the polypeptides of the invention with the known NMR (nuclear magnetic resonance) and x-ray crystal three-dimensional structures as templates. Table 5 shows: "PDB ID", the Protein DataBase (PDB) identifier given to template structure; "Chain ID", identifier of the subcomponent of the PDB template structure; "Compound Information", information of the PDB template structure and/or its subcomponents; "PDB Function Annotation" gives function of the PDB template as annotated by the PDB files (<http://www.rcsb.org/PDB/>); start and end amino acid position of the protein sequence aligned; PSI-BLAST score, the verify score, the SeqFold score, and the Potential(s) of Mean Force (PMF). The verify score is produced by GeneAtlas™ software (MSI), is based on Dr. Eisenberg's Profile-3D threading program developed in Dr. David Eisenberg's laboratory (US patent no. 5,436,850 and Luthy, Bowie, and Eisenberg, Nature, 356:83-85 (1992)) and a publication by R. Sanchez and A. Sali, Proc. Natl. Acad. Sci. USA, 95:13597-12502. The verify score produced by GeneAtlas normalizes the verify score for proteins with different lengths so that a unified cutoff can be used to select good models as follows:

Verify score (normalized) = (raw score – 1/2 high score)/(1/2 high score)

The PFM score, produced by GeneAtlas™ software (MSI), is a composite scoring function that depends in part on the compactness of the model, sequence identity in the alignment used to build the model, pairwise and surface mean force potentials (MFP). As given in Table 5, a verify score between 0 to 1.0, with 1 being the best, represents a good model. Similarly, a PMF score between 0 to 1.0, with 1 being the best, represents a good model. A SeqFold™ score of more than 50 is considered significant. A good model may also be determined by one of skill in the art based all the information in Table 5 taken in totality.

Table 7 correlates nucleotide sequences of the invention to a specific chromosomal location when assignable.

Table 6 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-971, their corresponding polypeptide sequences SEQ ID NO: 972-1942, their corresponding priority contig nucleotide sequences SEQ ID NO: 1943-2709, their corresponding priority contig polypeptide sequences SEQ ID NO: 2710-3476, and the US serial number of the priority application (all of which are herein incorporated in their entirety), in which the contig sequence was filed.

Table 8 is a correlation table of the novel polynucleotide sequences SEQ ID NO: 1-971, the novel polypeptide sequences SEQ ID NO: 972-1942, and the corresponding SEQ ID NO in which the sequence was filed in priority US application bearing serial number 60/324,631, filed September 24, 2001.

115
Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
adrenal gland	Clontech	ADR002	6-9 21 24 31 37 40 44 51 60 62-63 65-66 70-71 75 78 80 83 86 90 108 113 117 135-136 148 150 154-155 158-160 162 167 173-174 182 192 197 201-202 209 213 219-220 222 229 244 248 252 254-256 260 263 270 273 289 307 332 334 337 345-346 371 373 392 402 407-408 423 467 486 491 501 503-504 511 516 541 543 548 563-564 568 576-577 581 604 606 622 627 643-645 673 686 707 711 718 734 740-742 757 769 772 799 815 824 831 846 852 876-877 879 882-885 888 904 918 940 944-945 952 954 962 971
adult bladder	Invitrogen	BLD001	21 28 47 51 58 60 66 70 104 144 166 180 187 218 223 258 266 270-271 278 280 292 310 322 335-336 352-353 440 466 473 484 542 548 597-598 604 646 685 750 800 953 971
adult brain	BioChain	ABR012	110 428 740-742 928
adult brain	BioChain	ABR013	298
adult brain	Clontech	ABR001	20 59 61 84 89 94 128 138 142 153 164 180 201 279 281 287 290 297-298 301 307 313 324 380 406 414 432 451 495 571 575 705 757 798 800 831-832 861 863 876 883 944-945
adult brain	Clontech	ABR006	7-9 16-17 20 29 34-36 49-50 52 80 83-84 88 94 96 109 118 120 124 126-127 133 135-136 150 155 197 202 210 218 230-231 233 236 238 247 252 261 270 272 274 278 281-284 286 291-292 296 300 305 320 335-336 347 352 358 361 363 370-371 381 388-390 393-394 396 401 403 417 422 435 451 459 471 482 489-490 498 507 510 516 532-533 552 565 568 582 590 592-593 600-601 607 609 615 622 633-634 655 687 691 708 726 732 736 750 753-754 757-758 766-769 781 784 789-791 798 815 850 852 858-865 875 879-880 886-887 916 921 939 948 952 957 970
adult brain	Clontech	ABR008	7-10 12-14 16 21 23-24 29 33-35 38 41 44 48 51-52 54 61-62 65-66 68-71 78-80 83 85 89 96 100 105-110 113 115-116 118 122-123 126 131-133 135-136 138 143 146-149 151-155 158 161 164 172-174 180 182-183 186-189 192 196-197 201-202 205-206 213 215-216 220-221 223 228-229 231 236-238 246 248 251-252 255-256 258 260-261 264-265 269-271 274 276-277 284 287-289 291 296-298 301 303 306 309 311 313 319 326 328-329 331 333 337 348 352 357 360 363 365 367 370-372 378 381-382 384 386-390 394 397 399 401 406-407 410-412 414-416 420-422 426 441 443-444 449-452 455 459 466 468 475 479 489 491-494 496-498 505-515 528 533 546 548 550-557 563-568 582 592-594 603-607 610 612 615 622 627 633 636 641-645 648-649 657-665 677 685 689 695-696 705-706 711 718-723 726 729 732 734 736 740-742 744 757 768 770 774 776 784-787 789 798 800 808 812 818-819 827 829 850 856 862-865 868-871 873 875-877 879 882-883 885-887 890-891 895 906 909-910 915 927-930 933-934 939-942 944-945 947-948 955 957-958 970

Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
adult brain	Clontech	ABR011	894 939
adult brain	GIBCO	AB3001	5 21 37 52 61-62 70 88 96 110 112 122 127 133 135-136 138 142 145 150 154 164 204 218 281 305 312-317 353 466 502 581 648-649 685 691 726 757 831 852 864-865 886 916 918 970
adult brain	GIBCO	ABD003	6-9 21 24 40 43-44 48 50-51 58-59 63 70 78 83- 84 110 114-115 117 120 133 135-136 138 142 144 146 148 150 154-155 159-160 164 166 172 180 209 225 243-246 248 254 260 270 279 284 305 309 320-322 324-327 330 353-354 359 362 371 384-386 393 401-402 406 408 442 452 466-467 488-490 502 541 551 592-593 615 636 644 673 677 685-686 707 726 757 783 800 826 831 838 840 852 864-866 876-877 880 885 911 952-953 971
adult brain	Invitrogen	ABR014	38 70 149 197 210 296 352 444 604 615 640 705 833-835 864-865 882 915
adult brain	Invitrogen	ABR015	37 117 127 138 197 255-256 278 304 313 353 467 476-477 604 768 885
adult brain	Invitrogen	ABR016	1 14 70 117 134 138 142 279 309 353 377 466 911
adult brain	Invitrogen	ABT004	13 24 28 36 38 41 44 51 65 75 83 89 114-115 117 127 138 142 148 152 166 173-174 182 203 211-212 223 229 243 255-256 260 265 270-271 277-278 289-290 295 309 319 322 328 331 346 352 358 362 370-371 377 380-381 386 401 416 420 437 453 455-457 491 548 563-564 568 575 581 588 606 768 783 798 833-835 838 880 899 903 970
adult cervix	BioChain	CVX001	5-7 16-17 19 21 23 25 31 36 44 48-51 59 62-63 65 70 75-76 78 80 85 89-90 96 99-100 108-110 112 115 117-118 120 122 124 126 133 135-136 138 142 147 150 153-155 164-166 172 174 182 185-186 188 190 192 194 196 200-201 209 215-216 218 225 229 235 239-241 246 255-256 260-261 267-268 270 274 276 278 287 291 296-297 311 315 329-330 332 335 342 351-353 364-365 368 371-372 374 378 394 404-405 414 436 438-439 446 459 463 466 470 472 474 482 489 502-504 526 533 541 551 567 569 577 582 585-586 588 595-596 612 614 631 633 636 641-643 645 648-649 666-667 681 689 691 697 701 705 718 726 734 740-743 748-750 754 757 768 783 785-786 800 815 819 826-827 829 831 838 852 857 864-877 879-880 882 886 890 906-907 916-917 922-923 929-930 933-936 948 956 961 971
adult colon	Invitrogen	CLN001	15 28 37-38 43 63 91 96 113 122 134 176 191 209 260 270 336 341 344 401 427 450-451 467 470 476-477 563-564 612 618 663 670 705 750 755 847 883 907 937 951 971
adult heart	GIBCO	AHR001	6 8-9 13-14 20-22 28 30 36 40-41 45-47 49-50 52 58-59 62-63 65-66 70-71 75 78 80 84-85 93 96 101 109-111 113 118 126 141 148 150-151 153 155 165 182 185-188 202 210 214-221 224-226 235-236 239-240 243-244 246 252 254

Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			258-259 263-265 271 274 282-284 297 311 319 327 329 338 340 343-344 348 350 353 360 365 387 396 402-404 408 421 426 430 452 455 464- 467 483 490 499 502-504 536 538 541 551 569 577 602 615 624-625 644 662 673 677 691 705- 707 715-716 726 734 740-742 751 757 761 784 787 800 819 822 838 852 864-866 876-877 881 883 885-886 890-891 895 903-904 906 917 919-920 929-930 937 948 966 971
adult kidney	GIBCO	AKD001	6-7 11-12 16-17 19 23-24 28 30 36-37 45-46 48 51 55 58-60 62 66 70-75 78-80 85 88 96 101- 102 104 108-109 117-118 120 122 125-126 133 135-136 138 141 144 150 154-155 159-160 162 164 166 168-169 172 174 177 185 187-188 191 197 200-201 208-212 218 222-229 231-232 239-241 244 246 254 258 266-272 278 280 289 292-293 295 305 311 313 331-333 343-344 353 357 364-365 368 371 378 406 409-412 423 435 442-443 457 461 466-468 485-487 519 526 533 548 556-557 563-564 567 569 577 582 588 612 615 629 636-637 643-645 648-649 659 662-663 685-686 697 707 726 740-743 750 754 763 772 800 812 824-826 829 838 852 861 864-866 876 880 882-883 885-886 903 906-907 919-921 929-930 944-945 948 953 957 962 970-971
adult kidney	Invitrogen	AKT002	6-7 15-16 19 21 23 25 44 48 51 58-60 63 66 70 72-74 80 89 92-93 96 99-100 109 114 117 120 123 135-136 141-142 148 153 155 159-160 162 173-174 182 185 187 194 201 203 226 228 231 235 239-240 244 263 270 273 279 287 293-294 301 309 322 326 334-336 355 368 371-373 378 390 409 414 417 423 465-467 476-477 491-493 502 526 544 548 563-564 568 582 606 612 615 636 644 648-649 659 662 672-674 681 691 731 740-742 744-745 765 768 774 782 784-786 800 824 827 829 840 863-865 867 873 882 884-886 891 903-904 906 917-918 929-930 937 944-945 948 966 970-971
adult liver	Clontech	ALV003	2-4 11 40 59 72-74 91 148 184 277 293 312 467 548 768 840 882 886 919 944-945
adult liver	Invitrogen	ALV002	6-7 10-11 13 19 25 28 36 38 45-46 49 63 65 71- 75 77 88 96 114 134 153 155 170 180 184 187 200 202 219 232 236 248 255-256 269 272 277- 278 297 309 332 353 356 362 407 409-412 445 453 466 468 487 494 496 517 533 561 581 612 643 645 701 750 760 826 829 840 881 883 903 916 918 929-930 937 948
adult lung	GIBCO	ALG001	6 13 21 25 31 70 75 83 99 117 122 133 166 186 208 224 228 233 241 273 287 289 295 305 332 337-338 353 365 414 421 441 466-467 479 491 548 551 563-564 636 648-649 705 740-742 757 868-871 899 921
adult ovary	Invitrogen	AOV001	5-7 11 13 16-17 19-23 25 28 36-38 40-43 45-46 50-51 54 59-60 63 65-66 70-78 82-83 88-90 93 96 99 102 105 108-110 117-118 120 122-124 131 133-136 142 144 147-150 154-155 159-160

118
Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			164 166 169-170 172-174 176-178 180 182-185 188 191 196-197 200-201 204 209-213 215-216 218 222-223 225 228-229 231 235-236 239-241 243-244 246 254-258 260 263-264 269-270 277-280 287 291-292 295-297 300 304-305 307 312-313 315-317 323 327 329-332 340-347 352-353 357-358 365 368 371 376 378 388-397 401 404-407 410-412 414-415 417 421 423 437 440-442 444 451 459 461 466 468-475 479 483-484 490 492-493 502-504 510-511 519 533-534 542 546 548 551 559 563-564 568 577 588 592-593 597 602 612 615 622 626 631 635- 636 643 645 648-649 659 670-671 673 681 683 685 689 691 698 707 711 715-716 724 731 740- 744 748 754 757-758 768 771-772 783 787 798 800 812 815-816 826 831-832 838 840 844 852 864-866 868-871 873 875-876 879-880 883 885-887 903-904 906-908 911 916-918 921-923 928-930 933-935 944-945 948 953 955 957-958 961 965-966 971
adult placenta	Clontech	APL001	24 49 78 95 201 237 349 353 490 637 747 953
adult spleen	Clontech	SPLc01	34-35 44 48-49 52 83 99 108 110 118 122 124 126 150 153 155 161 186 194 197 201 206 215- 216 223 232-233 241 248 270 280 287-289 296-297 306 331-332 355 363 373 396 414-416 444 446 451 463 466 468 509-510 533 552 556- 557 561 565-566 576 601 606 634 636 656 710 720 736-739 773 798 827 864-865 875 879 883 899 916 919-920 940 952 955 970
adult spleen	GIBCO	ASP001	8-9 12 49 68-70 79 83 88 101 105 112 114 118 122 133 148 174 218 224 231 260 287 295 316- 317 337 341 344 352-353 398 421 423 441-442 466 479 491 518 551 563-564 604 705 754 772 829 910 955 970-971
adult testis	GIBCO	ATS001	20-21 30 51 54 58 70 72-74 83 96 101-102 107- 108 117 141 150 153 155 187 203 211-212 220 222-223 244 252 273 279 287 289 296 298 330 338 344 353 365 417 443 467 484 489 503-504 577 582 662 686 707 757 907-908 955 970-971
bone marrow	Clontech	BMD001	1-4 12 14 21 24 27 37 43-47 58 70 79-80 85 87- 88 90 95 99 110 115-126 131 141 155 157-167 169 172 185 192-193 195-197 200 205 215-216 225 227 231 233 237 244 262 273 277 284 295 298 301 313 329 331 344 353 357 365 371 378 401 405-408 410-415 423 437 442 446 466-467 474 479 490 503-504 541 548 563-564 567-568 583 585-586 588 605 622 635-636 638 645 655 662 671 673 683 697 706 718 726 740-743 754 757 783 825-826 831-832 850 852 864-866 906 908 910-911 948 953 955 957 971
bone marrow	Clontech	BMD007	1 70 856
bone marrow	GF	BMD002	1-7 14 16 20 24 26 29 38 41 43-44 48 52 55 60 65 70-71 75 78 80 82 85 88 90 96 107 112 114 117 119 122 124-126 133 140 148-149 154-155 158 161 164 166-167 172 174 186 188 193 196 200 205 211-212 215-216 220-221 225 230 233

Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			237 239-240 242 248 251-254 257 261 264-265 273-274 284 287 290 294-298 305 307 310-311 316-317 331 337 341 344 348 353 355 357 365 370 373 376 378 387 389 398 401 408-409 414- 416 421 423-424 427 441-442 446 455 463 466-468 479 488 491 496 501 509 516 521 531 533 539 541 543 546-548 550-551 558 563-564 569 571 576 583 588 595 601 606 615 622 627 651 673 706-707 709 711 720 724-725 727-728 738-739 754 757 761 773 776 778-779 783 787 791-793 800 815 820 826 829 831 846-847 850 858 861-865 872 879 886 895 899 906-907 909-910 933-935 941 944-945 952 959
bone marrow	NULL	STM001	41 70 82-83 248 261 332 430 840 882
cultured preadipocytes	Stratagene	ADP001	8-9 21 23 41 48-49 58 66 70-71 79 82 92 105 138 148 166 169 186 222-223 248 280 297 313 329 348 353 407 423 442 452 466 492-493 551 577 606 637-638 659 705 714 750 752 838 853 864-865 880 882 886 907 929-930
diaphragm	BioChain	DIA002	387 466-467
endothelial cells	Stratagene	EDT001	5-6 8-9 12 17 21 23 30 37-38 40-41 45-47 49- 51 58-60 63 70-71 77-80 83 85-86 88 93 96 101 108-109 114-115 117-118 122-123 133 139 148 150 153 159-160 165 169 178 182 185 187 207- 208 213 223 225 228 233 239-240 243-244 248 255-256 261 270 273 279-281 287 292 297 299-300 309-310 312-313 332 336 353 355 360 363 365 378 392 396-398 400 404 406-407 410-412 415 421 432-433 435 437 439 441-442 444 452 466-467 492-493 502-504 544 548 555 569 581 585-586 592-593 604 612 615 635-638 645 662 685 689 691 706-707 715-716 731 734 747-748 754 777 783-784 800 815 819 832 852 856 879-880 883 886 888 890 893 896 904 906 913 922-923 929-930 953 966 971
esophagus	BioChain	ESO002	761
fetal brain	Clontech	FBR001	21 84 146 177 191 198 294 296 312 355 362 480 510 630 659 740-742 764 769 781 784 886 906 953
fetal brain	Clontech	FBR004	23 49 60 135-136 148 268 272 278 281 298 384 390 426 462 482 634 655 887 899
fetal brain	Clontech	FBR006	5-9 13-14 16 19 21 33-35 38 41-44 47-48 50-52 54 60-63 65 70 75-78 82-83 85 89 100 107 110 114 116-117 123 130-132 135-136 138 144 146-147 152-153 159-160 164 169 179-180 187-188 190 197-198 201 213 215-216 225 228 231 233 238 242 248 251-253 255-256 258 260-261 264 266-271 273 276 284 286-287 289 297 301-302 309 311 313 318 321 329 337 341 343 346 348 353 365 370-371 373 375 378 383- 384 390 397 399 415-416 422 426-427 434 437 440 444 446 452 462-463 466-467 469 471 475 482 489 491 496-497 502-504 506 509 511 521 528 533 537-538 546 548-549 563-564 566 574 581 585-586 588 590 592-593 599-604 606-607 612 627-628 633 636 645 651 659 662 668 673

Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			677 689 705-706 709 711 715-716 729 732 740-742 744 750 754 758 762 768 772 777 779 783 789 795-800 812 818-819 826 829 832 843 856 867 872-873 875 877 879 882 887 890 895 903-904 909-910 919-920 922-923 927-928 940-942 957 970-971
fetal brain	Clontech	FBRs03	138 200 822 944-945
fetal brain	GIBCO	HFB001	5 7 16-17 21 23 38 40 43 47 49 58-59 68-71 75- 76 78-79 83-85 89 96 102 106 108-110 112 117-118 120 126-131 133 135-152 154-156 159-161 166 169 185 196 213 219 231 243 245- 246 254-256 260 263 270 273-274 279 303-305 313 324 341 352-353 359-366 371 381 397 406-407 415 423 454 461 467 483 490 537 548 551 569 574 585-586 604 606 636 639 645 648- 649 662 670 683 685 691 706 708 740-742 746 751-752 754 758 761 768 774 798 831-832 861 864-866 879 883 887 902 906 908 911 918 944- 945 948 953 958 970-971
fetal brain	Invitrogen	FBT002	13 21 41 60 65 70 79 101 114 130 152 155 169 178 180 185 187 197 229 243 245 270 273 277- 279 281 287 290 293 297 306 309 322 324 335 342 352 358 395 423 458-462 466 514 548-549 551 575 588 606 612 645 685 706 714 746 754 829 880 886-887
fetal heart	Invitrogen	FHR001	5-6 17 24 28 47-49 52 54 60 65-66 71 75-78 80 83 85 89 96 98 107 120 126 135-136 140 146- 148 153 158 163-166 168 170 173 186 209 213- 214 218 221 229 231 233 235-236 243 248 251 255-256 259-260 265 267-270 276-277 284 288 296 306-307 315-317 329 331 334 337 344 353 371 373 376 378 388 396 408 415 421 425-426 441 444 450 452 464 466 468 475 479 497 499 505 508 528 533 536 541-543 551 568 571 577 585-586 604-606 641-643 662 664 683 688 732 734 740-742 754 772 801-806 815 821-823 831-832 855 863-865 873 875-876 881 883-887 891 898 904 906 916 928-930 942 948 953 957 971
fetal kidney	Clontech	FKD001	59 65 70 83 118 164 169 194 213 225 235 260 278 282-283 300 331 336 353 416 457 479 542 714 740-742 754 893 917-918 953 971
fetal kidney	Clontech	FKD002	8-9 11 21 23 48 63 65 77-78 83 100 105 107 116 121 131 133 135-136 140 147 155 164 172 188 213 218-220 243 252 254 257 270 276 284 287 292 297 326 338 348 353 366 376 388 406 408-409 418 446 466-468 489 496 505 528 541 544 550 552 571 601 612 622 634 647 656 691 705 746-748 758 766-767 772 774 781 794 807 827 831 833-835 837 854 858 861 863 882 891 906 940 943 957
fetal kidney	Invitrogen	FKD007	70 563-564
fetal liver	Clontech	FLV002	1 18 40-41 47 49-50 77 103 113 116 131 202 223 255-256 445 583 673 760 789 808
fetal liver	Clontech	FLV004	1-5 8-10 14 52 58 63 67 70 75 79 85 95 100 116 118 122 124 135-136 148 158 166 171 174 188

Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			198 200 209 213 225 233 243 280 284 296 302 307 330 344 353 356 366 371 373 376 389 423 435 444-445 452 466-468 489-490 496 503-504 530 533 541 548 561 581 583 585-586 590 606- 607 627 633 636 648-649 655 662 690 695-696 738-739 760 774 805 807 809-811 815 824 831 838 857-858 862 879 886-887 899 910 917 953
fetal liver	Invitrogen	FLV001	1 11-12 24 38 41 43-44 70 72-74 76 88 107 110 112 170 184 203 229 262 278 297 362 365 376 407 432 442-443 445-448 467 496 548 577 581 800 883 899
fetal liver- spleen	Columbia University	FLS001	1-31 33-38 40-54 58-60 63-80 83-108 110-114 117 120 122 124 135-136 141 147-148 153 155 159-160 162 166 168-182 184-185 192 196 199-202 213 215-217 225-226 229 236-237 244 253 269 278-280 282-283 288 298 302 304 306 313 315 322-323 325 329-332 341 344 348 353-357 363 365 376 378 397 399 406-408 414-415 417 421 423 426 441-442 445 451-452 461 466-468 476-477 486 489-491 502-504 530 533-534 541-544 548 551 556-557 563-564 567-569 581 585-586 588-589 604 607 612-614 622 636-638 641-642 644-645 662-663 670 673 682 684-686 689 691 695-696 704-709 715-716 734 738-742 744 750 754 758 760-764 769-772 776 783-786 800 812 815 826-827 829 831 838 840 848 852 861 864-866 874-875 879-880 882-883 885-886 896 903-904 906 911 913 916 922-923 929-930 937 944-946 948-949 953-954 966 970-971
fetal liver- spleen	Columbia University	FLS002	2-6 8-11 13-14 16-17 20-21 24-29 31 36-37 40 43-44 47 50 52 54 59-60 63-65 67-80 86-90 94- 98 101-103 106-110 112-113 120 122-124 133 135-136 138-139 142 148 153 155 158-160 162 166 169-170 172 174 176 179-180 182 184 188 191-192 196-197 199 205 211-213 215-216 218-219 222 225-226 231 233 251 254 260 262 267 269-270 272 280 282-283 287 290 293 297 299 301 303-304 306-307 313 316-317 319 321-322 328-329 331-332 341-342 344 353 355-357 363 365-366 378 382 386 388-389 391 406-407 410-412 414-415 421 423 431-432 441 445 447-448 451-453 457 461 467 474-477 484 487 490 502 506 518 521 530 542-544 546 548 551 568-569 574 580 583-593 602 604-605 612-613 622 629 635-636 638 641-644 652 662 670-671 682-683 685 688-693 706-716 734 740-744 747-748 760 762 772 774 783-786 808 815-816 824 826-829 831-832 838 852 866 875 878 885 903-904 906-907 913 916 921 925-926 929-931 937 941 943-946 949 952 970
fetal liver- spleen	Columbia University	FLS003	1-5 10 16-17 30 33 60 70 72-74 76 79 88 96-97 99 108 113 117 122 124 126 153-155 169-170 172 179 184 186 191-193 199 223 229 239-240 260 290 297 302 313 332 344-345 350 353 363 376 387 402 406 445 447-448 476-477 486 516

122
Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			519 541-543 548 555 561 605 613-614 637 652 740-742 744 754 761-762 765 776-777 783-784 791 799 819 829-830 842 844 866 879 882 886 890 916 944-945
fetal lung	Clontech	FLG001	7 68-70 92 108 166 170 173 180 261 268 273 353 355 365 498-499 519 868-871 883 886 890 970
fetal lung	Clontech	FLG004	60
fetal lung	Invitrogen	FLG003	1 8-9 36-37 41 43 70 79 107 113 134 148 166 177 188 202 205 223 243 280 310 316-317 332 376 407 437 450-451 454 486 522 552 558 605- 607 631 685 731 761 772 781-782 863 868-871 899 921 941 951 965
fetal muscle	Invitrogen	FMS001	24 36 41 70 79 101 113 117 126 148 167 177 198 201 223 226 265 281 344 350 387 398 409 428 466 508 559 563-564 575 588 608-610 613 636 670 707 738-739 828 880 884 903 917 921 953 971
fetal muscle	Invitrogen	FMS002	5 13 18 21 23 36 41 44 47 52 63-64 70 87 98 100 105 110 126 135-136 147 153-154 166 185-186 213-214 218 220-221 231 236 259 265 271 280 284 286-287 306 316-317 329-330 338 341 361 371 373 388-389 394 396 408 421 423- 424 426 428 441 464 466-467 473 498 508 521 533 536 541 543 551 558 567 569 583 600 606 613 622 677 744 748 754 791 794 805 818 831 855 862 864-865 875 877 881-882 885 927 939 944-945 960
fetal skin	Invitrogen	FSK001	1 7 24-25 29 32 36 38 41 63 68-71 75 79 83 85 88 104-105 108 116-118 122 127 138 142 146 148-150 154 158 169-170 173 178 182 185 187 190 197-198 200 204 209 211-212 223 233-235 243 245 251-252 254 261 264 268 270 277 280 290 297-298 305 316-317 326 330-332 334 336 341-342 350-351 353 357 376 379 395 401 406-407 409-412 414 423-424 440 443-444 450-451 455 457 466 472 489 492-493 507 516 522 551-552 563-564 572 577 583 587 606-608 611-614 619 622 627 630 634 643 652 660 671 675-679 685 691 705-707 731 733 738-742 744 753 768 772 812 826 830 840 843 848 860 864- 865 868-871 883-885 899 903 906-907 910 916-917 928 948 953-954 956 965
fetal skin	Invitrogen	FSK002	1 5-6 23 36 38 42 44 48 54 65 68-69 78 83 85 94-95 98 109-111 114 116 122 135-136 138 147 149 155 164 166 173-174 178 182 186-187 198 201 206 210 213 220 223 225 229 231-232 237 245 249 255-256 262 268 270 284 288 292 296-298 302 306 321 337 346-347 350-351 365 367 370 388 396 403 406 409 423-424 433 439 450-452 472 479 486 489 492-494 506 528 533-534 541 543 546 548 552-553 556-558 566 571 577 583 608-609 622 627 643-644 658 675 705-706 720 738-742 748 754 763 773 776 791 812 822 826 830-831 837-843 854-855 861 864-865 875 879 882-883 885 891 895 899 904

Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			906-907 910 916 928 937 942 944-945 948 958
fetal spleen	BioChain	FSP001	6 24 352
fibroblast	Stratagene	LFB001	5-6 31 36-37 45-46 49 59 65 70 96-97 109 134 144 146 150 155 186 192 196 226 254 257 273- 274 279-280 327 353 371 397 407 410-413 415 417 466-467 541 612 629 643 673 685-686 688 691 697 704 776 815 817 826 829 864-865 879 911 919 942 948 953 965
induced neuron-cells	Stratagene	NTD001	15 19 23 48 59-60 70 89 96 105 113 120 128 133 137 146 148 163 197 213 236 245 295 315 331 353-355 371 395 451 466 500 541 551 644 674 681 685 695-696 698 715-716 763 807 854 866 940 944-945 948 958-959 966
infant brain	Columbia University	IB2002	5 7-9 13 21 24 28 36 38 40 47 55 58-61 63 65- 66 68-69 71 75 80 84-85 88 93 107 109-110 115 122 126 128 130-131 135-138 142 146 153 155 161 166 172 180 182 184 187 197 200 203- 204 211-213 218 220 222-223 225 228 245 260-261 263 270 274 278 280 282-283 290 296 303-307 309 316-317 320 322 331-332 334-335 358 362 365 367-370 378 382 384 389 401-402 406-407 416-417 423 437 442-443 450 455 459 462 471 479-483 489-490 519 548 554 563-564 569 575 578 583 592-593 599 612 625 639-640 659-660 685 691 706 731 740-742 750 754 768 784 791 800 819 825 829-830 862 866 879 882- 886 903 908 911 916 918 925-926 942 944-945 962 966 970-971
infant brain	Columbia University	IB2003	8-9 21 30 34-36 38 41 47 51 58-60 70 75 78-79 85-86 92-93 96 107 110 115-116 118 127-128 135-136 139 142 146 149-150 153 155 159-160 164 166 174 176 180 182 187 210 213 218 220 222 228 245 250 252 260 269-270 273-274 277 280-281 296 304 312-313 316-317 322 326-327 332 334 341 344 353 358 361-362 365 367-370 401 407 415-416 421 423 441-442 451 458 471 479 482-483 488-489 503-504 507-508 510 563-564 569 571 583 606 670 685 702-703 736 750 768 776 785-786 788-789 798 812 819 831 840 851-853 860 863 866 873 879 883-885 887 890 903 909 911 916 918 938 940-941 944-945 962 966 970
infant brain	Columbia University	IBM002	5 8-9 28 84 135-136 210 309 316-317 401 442 546 577 685 734 838 860 885
infant brain	Columbia University	IBS001	36 38 59-60 65 86 108 110 137 139 180 245 314 320 353 450 456 563-564 798 825 831 970
leukocyte	Clontech	LUC003	12 21 37 50 59 70-71 79 99 148 155 161 174 222 234 236 284 365 455 486 516 556-557 563- 564 588 637 726 740-742 887 905 907 910 941 971
leukocyte	GIBCO	LUC001	1 5 12-13 17 21 23-24 33 36-37 41 43-46 48-49 52 59 70-71 76 78-79 85 87-88 96 99-100 108 110 116 118-119 122 124 127 133-134 141 146-148 155 159-161 163-164 166-167 169 172-174 187-191 194 197-198 202-207 213 224-226 228-229 231 234 236 238-249 251-258

124
Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			260 263 271 280 284-289 295 298 304 309-311 313 315-317 335 341 353 357 359 365 371 387- 388 402 406 408-409 417 423 442-443 448 466-468 479 489-490 492-493 496 501 503-504 532 541 546 548 551 556-557 563-564 568-569 571 584 605-606 612 615 622 635-637 641-649 653 655 662 691 693 697 706-707 715-716 718 732 734 740-742 747-748 769 772 783-784 800 812 826 831 840 866 876 882-883 896 899 903- 904 906 909-910 921 925-926 933-936 938 940 947-948 953 955 971
lung tumor	Invitrogen	LGT002	6-7 12-13 20-21 23 25 29 36-37 41-42 45-48 51 55 63 65 71 78-79 85 96 99 110 113-114 117 122 139 148-149 153-155 159-160 162 166 169-170 174 182 184-185 187 192-194 200 202 204 209 219 224-227 232 235 237 252 255-256 268 270 277-278 290 295 297 299-300 304-305 308-310 316-317 322 329-331 335 341 344 346 353 365 371-372 375 380 386 392 394 396 400- 402 406-407 410-412 423 431 439 444 452 466-468 470 475-478 487-489 503-504 542 544 546 548 551-552 556-557 563-564 579 581 588 612 631 636 643-645 648-649 662 668-671 673 679 681-683 685 691 706 715-716 740-743 751-752 754 768 772 783 798 812 819 826 828 831 840 843 852 854 857 863 877 879 882-883 886 890-891 895 906-907 925-926 933-935 938 948-951 953 957 970-971
lymph node	Clontech	ALN001	25 44 60 70 76 96 125 138 187 315 340 353 397 446 467 489 707 740-742 879 931 944-945 971
lymphocyte	CA-46	DGD001	37 83 159-160 440 650 787 900
lymphocytes	ATCC	LPC001	1 5 12-13 23 25 31 33-37 42-44 48 54 58-59 71 76 78 85 89-90 99 116-117 119 122 124 147- 148 153 161 165-166 169 174 181-182 187 189 191 194-195 206-207 210 226-227 237 241 243 251-252 260-261 268 272 275 280 284 287-288 301 306 311 316-317 322 327 338 348 355 357- 358 365 371 377 381 389 407 422-423 433 450- 451 455 467 483 486 490 501 503-504 541-543 546 548 556-557 568-569 607 643-644 662-663 670 688 691 694-697 706 718 734 740-743 789 831 840 846-847 850 861-862 864-865 883 887 903-904 908 910 933-934 944-945 955 970
lymphocytes	RPMI-8226	DGD004	159-160
macrophage	Invitrogen	HMP001	21 44 63 112 119 153 190 211-212 222 225 248 255-256 365 388 401 406 450 459 463 630 644 705 826 882-883 886 956
mammary gland	Invitrogen	MMG001	5 7 12-13 19-21 25 29 36-38 41-42 44 47 51-52 59 65 68-71 76-77 79 86 89-90 101 104 107 110 112 116-117 122 127 131 135-136 138 142 148-149 153 166 169-170 172-174 177 180 182 184 191 204 211-212 223 225 235 243 248 252 256 260-261 267 270-271 276-277 280 296-297 302 305-306 309 313 316-317 331-332 335-336 340 344 351-352 365 371 377-383 387-388 390

Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			400 404 406 409-412 414 423 426 443-444 450 452 455 457 466 470 489 491-494 499 511 522- 524 548-549 555-557 563-564 568 575 577 581 588 592-593 606 612 636 643 645 659 670-671 677 685 706-707 711 718 724 730-731 734 740-742 744 748 757-758 761 766-768 772 783 785-786 800 812 816 824 829 838 840 848 852- 853 857 861 868-871 880 882-883 885-886 899 903 906 916 921 929-930 933-935 938 944-946 951 970-971
melanoma from-cell-line- ATCC-#CRL- 1424	Clontech	MEL004	5 17 21 31 47 71 78-79 99 112 114 117 122 124 131 159-160 170 172 201 215-216 219 243 316-317 353 422-423 466-467 494 533 543 567 570 585-586 707-708 812 826 831 858 863 879 893 895 916 919-920 952 955 959 970
mixture of 16 tissues/mRNA*	various vendors	SUP002	1 7 24 36 38 44-46 60 66 70 89 119 127 169 182 194 205 210 218 221 223 227 229-230 235 255-256 270 273 277 281-283 286 293 296 302 309 347 353 365 426 444 456 462 466-467 494 553 592-593 601 630 634 651 677 685 705 718 840 885 896-899 902 910 918 928 931 948 955
mixture of 16 tissues/mRNA	various vendors	SUP004	495
mixture of 16 tissues/mRNA	various vendors	SUP007	38
mixture of 16 tissues/mRNA	various vendors	SUP005	1 148 221 365 512 516 634 776 833-835 879 912-914 941
mixture of 16 tissues/mRNA	various vendors	SUP008	1 5 13 38 61 117 135-136 155 161 172 243 250 290 401 466 703 911
mixture of 16 tissues/mRNA	various vendors	SUP009	1 38 70 96 110 243 307 768 811 939 953 956 971
neuronal cells	Stratagene	NTU001	13 21 45-46 60 68-70 83 89 107 139 148 178 197 207 218 220 248 251 284 309 409 526 541 563-564 585-586 604 633 660 677 685 699-700 715-716 726 752 761 763 774 776 782 844 858 863 872 877 882 899 914 940 950
pituitary gland	Clontech	PIT004	28 31 59 68-69 79 85 89 98 101 108-109 115 127 134 150 159-160 165 169 198 218 231 235 260-261 273 297-298 313 322 327 330 353 364 367 395 402 459 466 541 568-569 574 641-642 670 691 734 757 831 883 906 911 916 919-920
placenta	Clontech	PLA003	5-6 20 47 55 59 63 71 83 85 99 108 116 118 124 135-136 138 140 153-154 158 161 171 174 182 225 243 270 279 284 287 297 302 334 349 353 389 391 393 409 444 447-448 467 488 490 496 539 581 592-593 601 622 634 656 709 720 738-742 765 787 799-800 812-814 820 824 829 831 839 844-845 861-863 878 882-884 910 916 919-920 942 948 952 958
placenta	Invitrogen	APL002	12 24 37 68-70 82 102 110 117 174 208 280 349-352 423 492-493 548 800 921
prostate	Clontech	PRT001	6 13 20 23 25 37 42 49-50 65 70-71 76 85 107 117 150 162 166 203 215-216 219 254 270 277 279 288 297 301 305 307-308 310 321 338 353 366 376 379 402 414 466-467 469 525-528 567 580 585-586 588 602 648-649 673 686 706 718

Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			726 735 761 768 787 820 829 831 852 866 879
rectum	Invitrogen	REC001	6-7 18 43 66 79 96 113 138 148 166 170 182 202 207 321-322 341 356 414 443-444 450 466 470 484 489 491-493 530 588 615-619 707 734 766-768 793 807 848 868-871 880 883 916 921 928 938
retinoic acid-induced-neuronal-cells	Stratagene	NTR001	5 21 23 25 52 62 70 83 96-97 120 123 148 154 180 194 271 306 310 314 389 489 524 541 572 606 686 690 710 726 800 862-863 882
saliva gland	Clontech	SALs03	66
salivary gland	Clontech	SAL001	5 28 30 60 63 65 70 79 107-108 117 126 130 133 135-136 144 148 167 172 196 208 213 270 280 287 294 326 353 365 380 457 467 533 575 620-623 635-636 648-649 715-716 787 826 838 868-871 904 956 965 971
skeletal muscle	Clontech	SKM001	43 48 66 75 79 117 126 146 155 167 197 201 214 217 219 221 225 247 254 259 265 297 301- 302 309 329 387 394 428 497 598 638 691 724 752 757 829 864-865 881 883 948
skeletal muscle	Clontech	SKM002	387
skeletal muscle	Clontech	SKMs03	387
skin fibroblast	ATCC	SFB001	89 387
small intestine	Clontech	SIN001	5-7 12 17-18 21 24-25 27 44-46 58-60 62-63 70-71 77-78 87-88 96 99 101 109 111 117 122- 124 135-136 140 148 150 161 166 172 174 176 186 191 194-195 201 209 213 221 223 225 228- 229 234 243 246 260 273 280 284 293 297 300- 302 306-308 310 313 326 331-332 340-343 347 353 357 362 365 371 378 393 404 416 432 445- 448 459 467 474 489 492-493 521 529-531 533 543 556-557 559 602 605 622 631 638 643 645- 646 668 670 697 726 740-743 757 761 763 768 772-775 780-782 784-787 812 829 831 852 854-857 876 879 885 896 904 906-908 910 921 931 944-945 948-949 953 958 960 970
spinal cord	Clontech	SPC001	8-9 12 44 48-49 51 65 70 75 108 113 116-117 120 122 127 135-136 150 155 159-161 164 166 174 182 185-186 188 219 234 260 273 276 278 286 290 292 301 309 326 337 347-348 353 362 365 378 405 420 429 451 469 474 479 491 499 532-537 542-543 563-564 636 645 648-649 673 683 686 691 708 718 720 726 740-742 758-759 768 772 784 787 800 824 829 832 873 885-887 911 944-945 953 957-958 962
stomach	Clontech	STO001	6-7 34-35 59 65 70 72-74 172 237 255-257 293 307 353 371 402 410-412 569 600 636 673 866 868-871 879 906 931 955 970
thalamus	Clontech	THA002	5 13 28 59 63 77 79 115 117 122 127-128 138 142 146 182 187 190 197 205 220 252 307 309 328 336 344 367 386 421 435 437 489-490 497 508 538-539 548 581 615 685 691 705 726 756- 757 759-760 768 826 879-880 883 886-887 896 902 921 958 971
thymus	Clontech	THM001	5-6 17 44 47 49-50 52 59 68-71 85 87-88 96 100 104 117-118 122 157 159-161 164 169

127
Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			173-174 179 186 197 209 215-216 220 225-226 254 273 280-281 284 288 298 306 310 331 344 348 352-353 364 406 414 432 442-443 466-467 475 479 489 492-493 498 528 541-542 552 556-557 559 567 571 576 580 604 606 612 628-630 648-649 673 693 697 734 743 757 760 768 776 784 800 824-826 832 848 861 864-865 886 904 907 921 941 953 962 970
thymus	Clontech	THMc02	5-6 12 14 19-21 23-24 37 43-44 47 49-50 54 60 65 70-71 85 87 95 99 108 110 116-117 119 122 124 129 131 140 146 148 154 161 166 174 179 193 197 201 204 206-207 213 220 222 235-236 242 251-253 255-256 258 260-262 266-268 270-272 278 281-288 296-298 300-301 326 341 344 347-348 351 357 359-360 365-366 376 378 388-389 393 406-408 415 439 444 447-448 466 468 472 475-477 479 482 489 492-493 502-504 521 533 543 546 548 550 556-558 560-564 568 576 585-586 588 591 606 622 636-637 643-644 650-656 670 673 709 718 738-742 758 773 776 784 789 796 812 827 829 832 838 849-850 863-865 876 886 899 903-904 906 910 916 919-921 929-930 940-941 948 952-953 958 970-971
thyroid gland	Clontech	THR001	5 7 12 14 20-21 23-24 28 30 42 45-46 48 50-51 58-59 61 65-66 70-71 76-80 83 89 93-94 96 98 102 104 107-109 112 117-118 124 126 129 134-136 140 144 147 153 159-161 165 167 172-173 184-187 192 194 196-197 201 208 215-216 218-220 223 225 235 246 248 253 260-261 266 268-270 273 280 284 292 297-298 302 304 310 315 321 330 334 336-337 344 348-349 351 353 365-366 378 387-388 392 401 404 406-408 410-412 424 429 437 444 466-467 469 472 475 479 485 487 491 500 519 533 540-544 548 551 572 581-583 602 606 609 622 636 641-645 648-649 659 673 681 697 702 705-706 711 715-716 740-743 754 757 764 769 772 787 789 815 824 828-830 848 852 856 864-866 868-873 875-876 879 882-883 885-891 902 904-906 908 918-923 929-931 940 944-945 948 956 962 966 971
trachea	Clontech	TRC001	7 48-50 59 75 79 85 117 122 153 166 173 185 220 243 267 270 296 371-372 406 414 419 429 432 467 492-493 501 569 573-574 582 697 734-735 744 796 819 829 868-871 876 879 890-892 903
umbilical cord	BioChain	FUC001	1 5-7 18 21 25 31 43 45-46 51 59 70 75 83 88-89 96 98 105 108-109 115 117-118 120 122 133 138 150 155 166-167 174 182 185 190-191 196 200-201 218 220-221 223 228-229 235 241 254 270 273 279 281 287 291 297 307 315 330 334 336 342 349 353 363 365 372 406 410-412 423 441 466-467 469 475 490-493 521 526 548 551-552 561 584 611 622 629-630 637 645 680-683 685 691 707 718 726 750 755-756 768 775 800 831 838 844 855 878-881 883-885 893

128

Table 1

Tissue Origin	Library/RNA Source	HYSEQ Library Name	SEQ ID NOS:
			904 907 921 929-930 953 965 971
uterus	Clontech	UTR001	17 19 21 62 70 96 118 149 165-166 184 186 188 235 248 264 279 300 353 371-372 423 466 469 492-493 502 545-546 588 602 629 631 638 648-649 673 751-752 829 832 864-865 891 893-894 917 948
young liver	GIBCO	ALV001	2-4 10 21 38 40 51 70 76 113 117 147 202 243 270-271 275-276 287 306 341 457 479 491 547 637 662 867 906 921 970

*The 16 tissue/mRNAs and their vendor sources are as follows: 1) Normal adult brain mRNA (Invitrogen), 2) Normal adult kidney mRNA (Invitrogen), 3) Normal fetal brain mRNA (Invitrogen), 4) Normal adult liver mRNA (Invitrogen), 5) Normal fetal kidney mRNA (Invitrogen), 6) Normal fetal liver mRNA (Invitrogen), 7) normal fetal skin mRNA (Invitrogen), 8) human adrenal gland mRNA (Clontech), 9) Human bone marrow mRNA (Clontech), 10) Human leukemia lymphoblastic mRNA (Clontech), 11) Human thymus mRNA (Clontech), 12) human lymph node mRNA (Clontech), 13) human so\spinal cord mRNA (Clontech), 14) human thyroid mRNA (Clontech), 15) human esophagus mRNA (BioChain), 16) human conceptional umbilical cord mRNA (BioChain).

129
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
972	gi38223	Pan troglodytes	alpha1-globin	395	100
972	gi38225	Pan troglodytes	alpha2-globin	395	100
972	gi14336681	Homo sapiens	16p13.3 sequence section 1 of 8.	395	100
973	gi1293093	Mus musculus	arginase	869	85
973	gi202979	Rattus norvegicus	arginase	853	84
973	gi202981	Rattus norvegicus	liver arginase (E.C. 3.5.3.1)	853	84
974	gi1197498	Homo sapiens	H.sapiens arginase gene exon 1 and flanking regions (EC 3.5.3.1) (and joined CDS).	1158	100
974	AAR05306	Homo sapiens	Human arginase.	1158	100
974	gi178995	Homo sapiens	Human liver arginase mRNA, complete cds.	1155	99
975	gi13529083	Homo sapiens	Similar to arginase, liver, clone MGC:12405 IMAGE:3934479, mRNA, complete cds.	778	84
975	gi2047344	Rattus norvegicus	arginase II	497	60
975	gi497232	Xenopus laevis	arginase 3	486	61
976	gi14211500	Homo sapiens	secretory protein SEC8 mRNA, complete cds.	4912	100
976	gi14042555	Homo sapiens	cDNA FLJ14782 fis, clone NT2RP4000524, highly similar to Mus musculus Sec8 mRNA.	4912	100
976	AAB93175	Homo sapiens	Human protein sequence SEQ ID NO:12114.	4912	100
977	gi11999092	Homo sapiens	TGF beta inducible nuclear protein TINP1 (TINP1) mRNA, complete cds.	599	91
977	gi14091475	Homo sapiens	hairy cell leukemia protein 1 mRNA, complete cds.	599	91
977	AAB39308	Homo sapiens	Human secreted protein sequence encoded by gene 48 SEQ ID NO:188.	599	91
978	gi159714	Nephila clavipes	dragline silk fibroin	205	25
978	gi4519619	Haliotis discus	collagen pro alpha-chain	146	28
978	gi7670050	Xenopus laevis	type I collagen alpha 1	160	26
979	gi1657752	Homo sapiens	Human FE65-like protein (hFE65L) mRNA, partial cds.	3648	96
979	AAAY13459	Homo sapiens	Amino acid sequence of human Fe65-like protein.	3648	96
979	gi13377732	Rattus norvegicus	FE65	1495	50
980	gi1657752	Homo sapiens	Human FE65-like protein (hFE65L) mRNA, partial cds.	2037	75
980	AAAY13459	Homo sapiens	Amino acid sequence of human Fe65-like protein.	2037	75
980	gi13377732	Rattus norvegicus	FE65	1520	46
981	gi458542	Homo sapiens	H.sapiens mRNA for orphan nuclear hormone receptor.	437	98
981	AAW32536	Homo sapiens	Constitutively active receptor-alpha.	437	98
981	AAW93902	Homo sapiens	Human CAR receptor protein.	437	98
982	gi15157720	Agrobacterium	AGR_C_4595p	208	26

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		tumefaciens			
982	gi4520316	Sphingomonas paucimobilis	OH-DDVA meta-cleavage compound hydrolase	202	25
982	gi8777583	Sphingomonas paucimobilis	4-oxalomesaconate hydratase	191	29
983	gi14043978	Homo sapiens	interferon stimulated gene (20kD), clone MGC:14105 IMAGE:4309120, mRNA, complete cds.	427	98
983	gi6759541	Homo sapiens	H.sapiens mRNA for gpISG20 protein.	427	98
983	gi2062680	Homo sapiens	Human HEM45 mRNA, complete cds.	427	98
984	gi12804251	Homo sapiens	clone MGC:2652 IMAGE:3535451, mRNA, complete cds.	500	100
984	AAG02107	Homo sapiens	Human secreted protein, SEQ ID NO: 6188.	297	98
984	AAB95667	Homo sapiens	Human protein sequence SEQ ID NO:18448.	71	28
985	gi7023926	Homo sapiens	cDNA FLJ11336 fis, clone PLACE1010661, weakly similar to TESTIS-SPECIFIC PROTEIN PBS13.	774	50
985	AAB93694	Homo sapiens	Human protein sequence SEQ ID NO:13262.	774	50
985	gi8468615	Homo sapiens	TCP11 (TCP11) mRNA, complete cds.	523	46
986	AAR74205	Homo sapiens	Human death associated protein DAP-2.	287	27
986	AAW71367	Homo sapiens	Death associated protein-2 (DAP-1, DAP-kinase).	287	27
986	gi2094873	Homo sapiens	H.sapiens DAP-kinase mRNA.	287	27
987	gi12584159	Homo sapiens	zinc finger protein 268 (ZNF268) mRNA, complete cds.	1723	54
987	gi1020145	Homo sapiens	Human DNA binding protein (HPF2) mRNA, complete cds.	1687	53
987	gi5441615	Canis familiaris	zinc finger protein	1493	58
988	gi895921	Xenopus laevis	nucleolar phosphoprotein	152	21
988	gi601931	Oryctolagus cuniculus	neurofilament-H	145	22
988	gi5901659	Caenorhabditis elegans	XNP-1	148	24
989	AAV70539	Homo sapiens	Human Factor 8 Homologue.	213	36
989	AAB64627	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 137.	144	29
989	AAB64628	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 138.	144	29
990	AAB56613	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1191.	1120	100
990	gi182642	Homo sapiens	Human rapamycin-binding protein (FKBP25) mRNA, complete cds.	1109	96
990	gi182626	Homo sapiens	Human rapamycin binding protein (FK506) mRNA, complete cds.	1061	100
991	gi12746410	Mus musculus	coenzyme A diphosphatase	357	60
991	gi10764850	Arabidopsis	F1K23.5	127	33

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		thaliana			
991	gi5824849	Caenorhabditis elegans	Y87G2A.14	100	31
992	gi14348900	Homo sapiens	heat shock protein mRNA, complete cds.	2122	90
992	gi517065	Homo sapiens	Human chaperonin protein (Tcp20) gene complete cds.	2122	90
992	AAB56638	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1216.	2122	90
993	gi3298472	Mus musculus	zinc finger protein	828	91
993	gi4884374	Homo sapiens	mRNA; cDNA DKFZp586B1222 (from clone DKFZp586B1222); partial cds.	194	42
993	gi1254984	Rattus norvegicus	AKAP95	194	42
994	gi14042531	Homo sapiens	cDNA FLJ14769 fis, clone NT2RP3004189, weakly similar to VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1.	3327	99
994	gi14424653	Homo sapiens	clone MGC:16757 IMAGE:4131625, mRNA, complete cds.	3327	99
994	AAB95791	Homo sapiens	Human protein sequence SEQ ID NO:18751.	3327	99
995	gi2736284	Mus musculus	Ldb1a	2224	99
995	gi2582522	Gallus gallus	neural src interacting protein, long form; NSIP long form	2206	98
995	gi5123791	Homo sapiens	partial NLI gene for Nuclear LIM interactor, exon 2-11.	2175	100
997	gi13560797	Homo sapiens	ubiquitin specific protease mRNA, complete cds.	6817	100
997	gi37334	Homo sapiens	H.sapiens mRNA for tre oncogene (clone 213).	4510	93
997	gi5360127	Homo sapiens	NY-REN-60 antigen mRNA, partial cds.	4422	99
998	AAB51702	Homo sapiens	Human secreted protein sequence encoded by gene 32 SEQ ID NO:142.	391	65
998	gi5070621	Homo sapiens	retrotransposon L1 insertion in X-linked retinitis pigmentosa locus, complete sequence.	364	42
998	gi339776	Homo sapiens	Human transposon L1.2.	364	42
999	gi15157181	Agrobacterium tumefaciens	AGR_C_3718p	742	51
999	gi14022240	Mesorhizobium loti	probable D-lactate dehydrogenase	734	50
999	gi15075160	Sinorhizobium meliloti	PUTATIVE D-LACTATE DEHYDROGENASE (CYTOCHROME) PROTEIN	713	50
1000	gi14603247	Homo sapiens	Similar to RIKEN cDNA 5730409G15 gene, clone MGC:19636 IMAGE:2822323, mRNA, complete cds.	368	91
1000	AAB36613	Homo sapiens	Human FLEXHT-35 protein	368	91

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			sequence SEQ ID NO:35.		
1000	gi7022832	Homo sapiens	cDNA FLJ10661 fis, clone NT2RP2006106.	220	84
1001	gi4884142	Homo sapiens	mRNA; cDNA DKFZp586P2219 (from clone DKFZp586P2219).	607	99
1001	gi13421401	Caulobacter crescentus	MutT/nudix family protein	133	40
1001	gi13276687	Homo sapiens	mRNA; cDNA DKFZp761I172 (from clone DKFZp761I172); complete cds.	115	31
1002	gi4689229	Rattus norvegicus	b-tomosyn isoform	2075	94
1002	gi3790389	Rattus norvegicus	m-tomosyn	1976	95
1002	gi4689231	Rattus norvegicus	s-tomosyn isoform	1896	94
1003	AAB29645	Homo sapiens	Human membrane-associated protein HUMAP-2.	311	83
1003	gi1196425	Homo sapiens	Human endogenous retrovirus ERV3, pol-env-3'LTR region.	309	71
1003	AAB95392	Homo sapiens	Human protein sequence SEQ ID NO:17743.	273	77
1004	AAY21631	Homo sapiens	Ligand binding domain of nuclear receptor hTRbeta.	524	100
1004	gi31207	Homo sapiens	Human c-erb-A mRNA for thyroid hormone receptor.	497	100
1004	gi180253	Homo sapiens	Human c-erbA mRNA, complete cds.	497	100
1005	gi13097597	Homo sapiens	clone IMAGE:3542589, mRNA, partial cds.	2967	100
1005	gi2370155	Homo sapiens	Homo Sapiens mRNA for spinocerebellar ataxia 7.	1031	37
1005	gi3192954	Homo sapiens	ataxin-7 (SCA7) mRNA, complete cds.	1029	37
1006	gi13097597	Homo sapiens	clone IMAGE:3542589, mRNA, partial cds.	1853	99
1006	gi3192954	Homo sapiens	ataxin-7 (SCA7) mRNA, complete cds.	1270	40
1006	gi2370155	Homo sapiens	Homo Sapiens mRNA for spinocerebellar ataxia 7.	1268	40
1007	gi5924306	Plasmodium falciparum	ADA2-like protein	410	30
1007	gi170991	Saccharomyces cerevisiae	ADA2	396	31
1007	gi927705	Saccharomyces cerevisiae	Ada2p: probable transcriptional adaptor; YDR448W; CAI: 0.12	396	31
1008	gi1054747	Homo sapiens	H.sapiens DMA, DMB, HLA-Z1, IPP2, LMP2, TAP1, LMP7, TAP2, DOB, DQB2 and RING8, 9, 13 and 14 genes.	952	100
1008	gi38482	Homo sapiens	H.sapiens gene for major histocompatibility complex encoded proteasome subunit LMP7.	952	100
1008	gi596140	Homo sapiens	Human proteasome subunit LMP7 (allele LMP7B) mRNA, complete	952	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			cds.		
1009	gi9309471	Homo sapiens	DNMT1 associated protein-1 (DMAP1) mRNA, complete cds.	2385	96
1009	gi10432820	Homo sapiens	cDNA FLJ11543 fis, clone HEMBA1002816.	2385	96
1009	gi12804007	Homo sapiens	clone MGC:3485 IMAGE:3635724, mRNA, complete cds.	2385	96
1010	gi6942267	Danio rerio	bHLH transcription factor Mesp-b	80	27
1010	AAB95031	Homo sapiens	Human protein sequence SEQ ID NO:16776.	71	26
1010	AAY87132	Homo sapiens	Human secreted protein sequence SEQ ID NO:171.	70	51
1011	gi5815353	Homo sapiens	J domain containing protein 1 isoform a (JDP1) mRNA, complete cds.	782	98
1011	gi6453811	Mus musculus	J domain protein 1	548	72
1011	gi14192847	Rattus norvegicus	J domain protein 1	527	69
1012	gi5777952	Homo sapiens	nuclear prelamin A recognition factor mRNA, complete cds.	1946	99
1012	gi7021904	FE	cDNA FLJ10067 fis, clone HEMBA1001526, weakly similar to PERIPLASMIC HYDROGENASE 1 (EC 1.18.99.1). [Homo	1946	99
1012	AAB97260	Homo sapiens	Human NADP hydrogenase subunit 50.	1946	99
1013	gi5917666	Zea mays	extensin-like protein	472	27
1013	gi15145797	Sus scrofa	basic proline-rich protein	415	29
1013	gi6523547	Volvox carteri f. nagariensis	hydroxyproline-rich glycoprotein DZ-HRGP	376	32
1014	gi407468	Mus musculus	SEB4	1172	93
1014	gi407419	Homo sapiens	H.sapiens seb4D mRNA.	1141	96
1014	gi407421	Homo sapiens	H.sapiens seb4B mRNA.	1074	96
1015	gi12805043	Homo sapiens	clone IMAGE:3461487, mRNA, partial cds.	2281	100
1015	AAB75532	Homo sapiens	Human secreted protein sequence encoded by gene 27 SEQ ID NO:86.	680	92
1015	gi9802536	Arabidopsis thaliana	F17L21.25	247	27
1016	gi7020516	Homo sapiens	cDNA FLJ20424 fis, clone KAT02627.	1054	99
1016	gi10438720	Homo sapiens	cDNA: FLJ22363 fis, clone HRC06574.	1054	99
1016	AAB43450	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:895.	779	98
1017	gi10438720	Homo sapiens	cDNA: FLJ22363 fis, clone HRC06574.	1215	100
1017	gi7020516	Homo sapiens	cDNA FLJ20424 fis, clone KAT02627.	1212	99
1017	AAB43450	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:895.	644	97
1018	gi12653673	Homo sapiens	tubulin, gamma 1, clone MGC:1593 IMAGE:3345973, mRNA, complete cds.	2320	97
1018	gi183703	Homo sapiens	Human gamma-tubulin mRNA,	2311	96

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			complete cds.		
1018	gi4115724	Rattus norvegicus	tubulin	2295	96
1019	gi12653689	Homo sapiens	Similar to aspartyl-tRNA synthetase, clone MGC:1562 IMAGE:3344322, mRNA, complete cds.	2586	100
1019	gi179102	Homo sapiens	Human aspartyl-tRNA synthetase alpha-2 subunit mRNA, complete cds.	2548	98
1019	gi14250408	Mus musculus	Similar to aspartyl-tRNA synthetase	2521	96
1020	AAB63116	Homo sapiens	Human secreted protein sequence encoded by gene 39 SEQ ID NO:126.	2035	99
1020	gi14039848	Homo sapiens	testes development-related NYD-SP19 mRNA, complete cds.	1919	99
1020	gi10437135	Homo sapiens	cDNA: FLJ21108 fis, clone CAS05257.	1144	99
1021	gi6015493	Candida albicans	proline transport helper PTH1	345	39
1022	gi10434528	Homo sapiens	cDNA FLJ12816 fis, clone NT2RP2002609, weakly similar to 2-HYDROXYMUCONIC SEMIALDEHYDE HYDROLASE (EC 3.1.1.-).	1819	99
1022	AAB94324	Homo sapiens	Human protein sequence SEQ ID NO:14807.	1819	99
1022	gi4929585	Homo sapiens	CGI-58 protein mRNA, complete cds.	1023	54
1023	gi12803055	Homo sapiens	clone MGC:4614 IMAGE:3504258, mRNA, complete cds.	684	100
1023	gi8978394	Chlamydophila pneumoniae J138	Pmp_5	69	25
1023	AAAY39473	Homo sapiens	DNAX interferon-like receptor subunit 2 protein sequence.	73	33
1024	gi10433115	Homo sapiens	cDNA FLJ11783 fis, clone HEMBA1006005.	719	100
1024	AAB93824	Homo sapiens	Human protein sequence SEQ ID NO:13634.	719	100
1024	gi170454	Lycopersicon esculentum	cell wall hydroxyproline-rich glycoprotein	70	34
1025	gi14042550	Homo sapiens	cDNA FLJ14779 fis, clone NT2RP4000398, moderately similar to ZINC FINGER PROTEIN 140.	1372	58
1025	AAB93164	Homo sapiens	Human protein sequence SEQ ID NO:12091.	1372	58
1025	gi13560888	Homo sapiens	EZF1T-related protein 1 mRNA, complete cds.	1298	50
1026	gi11386005	Homo sapiens	hepatocellular carcinoma-associated protein HCA10 mRNA, complete cds.	2802	100
1026	gi13021980	Homo sapiens	hepatocellular carcinoma-associated protein HCA11 mRNA, complete cds.	2802	100
1026	gi13623377	Homo sapiens	clone MGC:10536	2802	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			IMAGE:3958645, mRNA, complete cds.		
1027	gi6063527	Homo sapiens	COPS7a mRNA for cop9 complex subunit 7a, complete cds.	1241	96
1027	gi6606550	Homo sapiens	COP9 complex subunit 7a mRNA, complete cds.	1241	96
1027	gi10433901	Homo sapiens	cDNA FLJ12426 fis, clone MAMMA1003113, highly similar to Mus musculus COP9 complex subunit 7a (COPS7a) mRNA.	1241	96
1028	AAV35909	Homo sapiens	Extended human secreted protein sequence, SEQ ID NO. 158.	854	77
1028	gi6063527	Homo sapiens	COPS7a mRNA for cop9 complex subunit 7a, complete cds.	622	80
1028	gi6606550	Homo sapiens	COP9 complex subunit 7a mRNA, complete cds.	622	80
1029	gi15030144	Mus musculus	Similar to RIKEN cDNA 1700127B04 gene	616	88
1029	gi4512613	Arabidopsis thaliana	F28K20.12	132	39
1029	gi15073552	Sinorhizobium meliloti	HYPOTHETICAL PROTEIN	72	48
1030	gi14290488	Homo sapiens	clone MGC:17110 IMAGE:4186365, mRNA, complete cds.	1787	100
1030	gi1381027	Homo sapiens	Human mRNA for phosphoribosylpyrophosphate synthetase-associated protein 39, complete cds.	1775	99
1030	gi436779	Rattus norvegicus	phosphoribosylpyrophosphate synthetase-associated protein (39 kDa)	1768	98
1031	gi15126730	Homo sapiens	clone MGC:16794 IMAGE:3866664, mRNA, complete cds.	1672	100
1031	gi10438924	Homo sapiens	cDNA: FLJ22509 fis, clone HRC11803.	1663	99
1031	AAB35408	Homo sapiens	Human 07CG27 gene protein.	412	29
1032	AAB95299	Homo sapiens	Human protein sequence SEQ ID NO:17530.	2238	99
1032	AAB67054	Homo sapiens	Human immune response molecule (IMUN) protein SEQ ID NO: 8.	1417	98
1032	gi2232116	Mus musculus	mutated immunoglobulin heavy chain	65	31
1033	AAV70242	Homo sapiens	Human RNA-associated protein-23 (RNAAP-23).	1595	97
1033	gi3037013	Homo sapiens	RRM RNA binding protein Gry-rbp (GRY-RBP) mRNA, complete cds.	1595	97
1033	gi3694986	Mus musculus	RRM RNA binding protein GRYP-RBP	1564	95
1034	gi13436062	Homo sapiens	zinc finger protein 144 (Mel-18), clone MGC:10336 IMAGE:3841545, mRNA, complete cds.	1805	97
1034	gi285933	Homo sapiens	Human mRNA for Mel-18 protein,	1805	97

136
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			complete cds.		
1034	gi6012977	Felis catus	Bmi-1 protein	1057	61
1035	gi3335159	Homo sapiens	bestrophin (VMD2) mRNA, alternatively spliced product, complete cds.	645	59
1035	gi3511242	Homo sapiens	vitelliform macular dystrophy protein (VMD2) mRNA, complete cds.	645	59
1035	gi3598876	Homo sapiens	vitelliform macular dystrophy protein (VMD2) gene, exon 11 and complete cds.	645	59
1036	gi13543935	Homo sapiens	clone MGC:12966 IMAGE:3139887, mRNA, complete cds.	227	69
1036	gi14250415	Homo sapiens	clone IMAGE:3868989, mRNA, partial cds.	213	66
1036	gi1006682	Neisseria meningitidis	outer membrane protein P64k or PM-6	54	33
1037	AAAY48501	Homo sapiens	Human breast tumour-associated protein 46.	987	98
1037	gi6635999	Homo sapiens	cardiovascular heat shock protein mRNA, complete cds.	869	100
1037	gi13623439	Homo sapiens	heat shock 27kD protein family, member 7 (cardiovascular), clone MGC:12642 IMAGE:4131506, mRNA, complete cds.	869	100
1038	AAG01992	Homo sapiens	Human secreted protein, SEQ ID NO: 6073.	296	100
1038	gi5911915	Homo sapiens	mRNA; cDNA DKFZp586M0622 (from clone DKFZp586M0622); partial cds.	81	32
1038	gi1181346	Paramecium bursaria Chlorella virus 1	a183L	66	34
1039	gi12655013	Homo sapiens	clone MGC:3036 IMAGE:3163787, mRNA, complete cds.	1353	100
1039	gi10440510	Homo sapiens	mRNA for FLJ00101 protein, partial cds.	141	35
1039	gi8132056	Trypanosoma brucei	phosphatase	137	30
1040	gi12655013	Homo sapiens	clone MGC:3036 IMAGE:3163787, mRNA, complete cds.	933	100
1040	gi10440510	Homo sapiens	mRNA for FLJ00101 protein, partial cds.	141	35
1040	gi8132056	Trypanosoma brucei	phosphatase	137	30
1041	gi12652649	Homo sapiens	ribosomal protein L28, clone MGC:3230 IMAGE:3504015, mRNA, complete cds.	557	98
1041	gi14603452	Homo sapiens	clone MGC:20081 IMAGE:4054251, mRNA, complete cds.	557	98
1041	gi15079503	Homo sapiens	clone MGC:20378 IMAGE:4561118, mRNA, complete cds.	557	98

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1042	AAW95630	Homo sapiens	Homo sapiens secreted protein gene clone gn114 1.	358	100
1042	AAG02785	Homo sapiens	Human secreted protein, SEQ ID NO: 6866.	245	98
1042	gi467292	Dictyostelium discoideum	glutamine-asparagine rich protein	231	24
1043	gi306809	Homo sapiens	Human glutathione S-transferase 2 (GST) mRNA, complete cds.	859	99
1043	gi306810	Homo sapiens	Human glutathione S-transferase (GST) Ha subunit 1 mRNA, complete cds.	859	99
1043	gi306815	Homo sapiens	Human glutathione S-transferase (GST) a-subunit mRNA, complete cds.	859	99
1044	gi825605	Homo sapiens	H.sapiens GSTalpha gene for glutathione S-transferase exon 2.	534	96
1044	gi12804085	Homo sapiens	Similar to glutathione S-transferase A2, clone MGC:10525 IMAGE:3943417, mRNA, complete cds.	534	96
1044	gi306811	Homo sapiens	Human glutathione S-transferase Ha subunit 2 (GST) mRNA, complete cds.	534	96
1045	gi306809	Homo sapiens	Human glutathione S-transferase 2 (GST) mRNA, complete cds.	702	100
1045	gi306810	Homo sapiens	Human glutathione S-transferase (GST) Ha subunit 1 mRNA, complete cds.	702	100
1045	gi306815	Homo sapiens	Human glutathione S-transferase (GST) a-subunit mRNA, complete cds.	702	100
1046	gi10438279	Homo sapiens	cDNA: FLJ22029 fis, clone HEP08661.	1071	100
1046	gi13097624	Homo sapiens	clone IMAGE:3608084, mRNA, partial cds.	615	60
1046	gi13325154	Homo sapiens	clone IMAGE:3635709, mRNA, partial cds.	609	47
1047	gi1755138	Homo sapiens	sorbitol dehydrogenase (SORD) gene, exons 8 and 9, and complete cds.	1228	100
1047	gi520450	Homo sapiens	sorbitol dehydrogenase gene, complete cds.	1228	100
1047	gi496078	Homo sapiens	Human L-iditol-2 dehydrogenase mRNA, complete cds.	1221	99
1048	AAB58331	Homo sapiens	Lung cancer associated polypeptide sequence SEQ ID 669.	478	97
1048	AAG01769	Homo sapiens	Human secreted protein, SEQ ID NO: 5850.	61	39
1049	gi13276623	Homo sapiens	mRNA; cDNA DKFZp761B1514 (from clone DKFZp761B1514).	1633	100
1049	gi14042525	Homo sapiens	cDNA FLJ14766 fis, clone NT2RP3004041.	1626	99
1049	AAB95783	Homo sapiens	Human protein sequence SEQ ID NO:18735.	1626	99
1050	gi15011995	Homo sapiens	clone MGC:13350	943	95

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			IMAGE:4333615, mRNA, complete cds.		
1050	AAB56131	Homo sapiens	Human secreted protein sequence encoded by gene 55 SEQ ID NO:225.	943	95
1050	gi12052900	Homo sapiens	mRNA; cDNA DKFZp564G0678 (from clone DKFZp564G0678); complete cds.	938	95
1051	gi10435735	Homo sapiens	cDNA FLJ13657 fis, clone PLACE1011563.	1819	99
1051	AAB94709	Homo sapiens	Human protein sequence SEQ ID NO:15705.	1819	99
1051	gi12043689	Plasmodium falciparum	merozoite surface protein	115	30
1052	gi12654509	Homo sapiens	Similar to SH3 and PX domain-containing protein SH3PX1, clone IMAGE:2958415, mRNA, partial cds.	259	42
1052	gi5410249	Homo sapiens	SDP1 protein mRNA, complete cds.	259	42
1052	gi4689258	Homo sapiens	sorting nexin 9 (SNX9) mRNA, complete cds.	259	42
1053	gi15074372	Sinorhizobium meliloti	HYPOTHETICAL TRANSMEMBRANE PROTEIN	102	39
1053	gi6002569	Coturnix japonica	POU-box protein brain-2	95	38
1053	gi13623799	Homo sapiens	FZD8 mRNA for seven-transmembrane receptor Frizzled-8, complete cds.	96	40
1054	gi28384	Homo sapiens	H.sapiens ADE2H1 mRNA showing homologies to SAICAR synthetase and AIR carboxylase of the purine pathway (EC 6.3.2.6, EC 4.1.1.21).	1372	99
1054	gi976252	Rattus norvegicus	AIR carboxylase-SAICAR synthetase	1334	95
1054	gi2111194	Gallus gallus	5-aminoimidazole ribonucleotide (AIR) carboxylase-5-aminoimidazole-4-N-succinocarboxamide ribonucleotide (SAICAR) synthetase	1248	88
1055	AAG01499	Homo sapiens	Human secreted protein, SEQ ID NO: 5580.	293	100
1055	gi9188555	Chlorella vulgaris	Homolog of chick heat shock factor 3	60	43
1055	gi3098309	Crithidia fasciculata	H1 histone-like protein p21	74	26
1056	gi3115347	Homo sapiens	BAC clone CTA-276O3 from 7q22-q31.1, complete sequence.	121	29
1056	gi1770454	Homo sapiens	H.sapiens mRNA for M-phase phosphoprotein, mpp11.	121	29
1056	gi313202	Drosophila hydei	mst101(2)	123	24
1057	gi14550467	Homo sapiens	spermatogenesis associated 2, clone MGC:15312 IMAGE:3953522, mRNA, complete cds.	184	26
1057	AAB95348	Homo sapiens	Human protein sequence SEQ ID NO:17635.	184	26

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1057	gi4730927	Homo sapiens	spermatogenesis associated PD1 mRNA, complete cds.	181	26
	gi13544063	Homo sapiens	clone MGC:13219 IMAGE:3959259, mRNA, complete cds.	254	54
	AAB28631	Homo sapiens	Human breast tumour-specific antigen B21GT2.	220	49
	gi8272464	Homo sapiens	human endogenous retrovirus W gagC3.37 G gag (gag) gene, complete cds.	202	46
1058	gi14334374	Homo sapiens	leucine zipper protein AF5alpha mRNA, complete cds.	649	74
1058	gi13529293	Homo sapiens	Similar to leucine zipper protein FKSG14, clone MGC:12540 IMAGE:3839409, mRNA, complete cds.	649	74
1058	gi14250169	Homo sapiens	Similar to leucine zipper protein FKSG14, clone MGC:14847 IMAGE:3511065, mRNA, complete cds.	649	74
1059	AAB62180	Homo sapiens	Human p95 protein.	2395	89
1059	gi15080468	Homo sapiens	Similar to RIKEN cDNA 1110018J12 gene, clone IMAGE:3865164, mRNA, partial cds.	660	100
1059	gi165520	Oryctolagus cuniculus	smooth muscle myosin heavy chain	144	24
1060	gi13436353	Homo sapiens	clone IMAGE:3632533, mRNA, partial cds.	166	35
1060	gi13186114	Homo sapiens	mRNA for rab interacting lysosomal protein (RILP gene).	166	35
1060	AAG02093	Homo sapiens	Human secreted protein, SEQ ID NO: 6174.	140	42
1061	gi10440398	Homo sapiens	mRNA for FLJ00032 protein, partial cds.	2228	60
1061	gi13752754	Homo sapiens	zinc finger 1111 mRNA, complete cds.	2168	59
1061	gi186774	Homo sapiens	Human Kruppel related zinc finger protein (HTF10) mRNA, complete cds.	2130	52
1062	AAAY17227	Homo sapiens	Human secreted protein (clone ya1-1).	1476	99
1062	gi12831176	Agelaius phoeniceus	gamma filamin protein	183	26
1062	AAAY94938	Homo sapiens	Human secreted protein clone ye78_1 protein sequence SEQ ID NO:82.	169	25
1063	gi14042641	Homo sapiens	cDNA FLJ14834 fis, clone OVARC1001270.	560	100
1063	AAB94558	Homo sapiens	Human protein sequence SEQ ID NO:15328.	560	100
1063	gi246315	Mus sp.	immunoglobulin lambda light chain variable region	58	39
1064	gi13752754	Homo sapiens	zinc finger 1111 mRNA, complete cds.	2304	58

140
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1064	gi10440398	Homo sapiens	mRNA for FLJ00032 protein, partial cds.	2294	59
1064	gi10436789	Homo sapiens	cDNA FLJ14345 fis, clone THYRO1001189, weakly similar to ZINC FINGER PROTEIN 91.	2081	59
1065	gi2226004	Homo sapiens	Human Tigger1 transposable element, complete consensus sequence.	338	39
1065	gi14349368	Homo sapiens	clone MGC:16385 IMAGE:3937996, mRNA, complete cds.	237	34
1065	AAB39379	Homo sapiens	Human secreted protein sequence encoded by gene 21 SEQ ID NO:128.	200	38
1066	gi2645086	Hydra vulgaris	collagen	86	39
1066	gi555157	Human herpesvirus 4	nuclear antigen 1	68	39
1066	AAW99376	Homo sapiens	Human fibroblast growth factor 2 24 kD isoform N-terminus.	67	33
1067	gi7453641	Drosophila melanogaster	casein kinase II beta subunit	53	37
1068	gi7022973	Homo sapiens	cDNA FLJ10749 fis, clone NT2RP3001915.	210	32
1068	AAB93129	Homo sapiens	Human protein sequence SEQ ID NO:12017.	210	32
1068	gi12053201	Homo sapiens	mRNA; cDNA DKFZp434A1031 (from clone DKFZp434A1031); complete cds.	210	32
1069	gi14042847	Homo sapiens	cDNA FLJ14957 fis, clone PLACE4000009, weakly similar to MYOSIN HEAVY CHAIN, NONMUSCLE TYPE B.	2482	99
1069	AAB95546	Homo sapiens	Human protein sequence SEQ ID NO:18167.	2482	99
1069	gi8308176	Homo sapiens	cingulin mRNA, complete cds.	1021	44
1070	AAB47134	Homo sapiens	CDIFF-15, Incyte ID No. 3478571CD1.	3992	99
1070	gi7022185	Homo sapiens	cDNA FLJ10260 fis, clone HEMBB1000973, moderately similar to Mus musculus schlafen3 mRNA.	1182	46
1070	AAB92636	Homo sapiens	Human protein sequence SEQ ID NO:10951.	1182	46
1071	gi172612	Saccharomyces cerevisiae	SSD1 protein	172	36
1071	gi172697	Saccharomyces cerevisiae	SRK1	172	36
1071	gi1230657	Saccharomyces cerevisiae	Ssd1p	172	36
1072	gi10438029	Homo sapiens	cDNA: FLJ21841 fis, clone HEP01831.	2194	99
1072	gi35019	Homo sapiens	H.sapiens nestin gene.	1642	78
1072	AAR60127	Homo sapiens	Human nestin protein is useful to identify brain tumours.	1642	78
1073	AAAY91424	Homo sapiens	Human secreted protein sequence	997	96

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			encoded by gene 12 SEQ ID NO:145.		
1073	AAY41495	Homo sapiens	Fragment of human secreted protein encoded by gene 70.	105	33
1073	gi3879551	Caenorhabditis elegans	contains similarity to Pfam domain: PF01391 (Collagen triple helix repeat (20 copies)), Score=56.4, E-value=2e-13, N=2; PF01484 (Nematode cuticle collagen N-terminal domain), Score=87.2, E-value=1.1e-22, N=1	108	29
1074	gi205278	Rattus norvegicus	male germ cell-associated kinase (mak)	2188	80
1074	gi53914	Mus musculus	rck	2159	79
1074	gi12002678	Homo sapiens	MAK-related kinase (LCK2) mRNA, complete cds.	1152	47
1075	AAG03354	Homo sapiens	Human secreted protein, SEQ ID NO: 7435.	373	100
1075	gi2511438	Homo sapiens	calcium/calmodulin-dependent protein kinase II mRNA, partial cds.	71	30
1075	AAB95290	Homo sapiens	Human protein sequence SEQ ID NO:17511.	66	42
1076	gi7578595	Mus musculus	teashirt 2	4578	89
1076	gi7527470	Mus musculus	zinc finger protein	1603	45
1076	gi3170196	Homo sapiens	antigen NY-CO-33 (NY-CO-33) mRNA, complete cds.	1636	52
1077	gi1403028	Calotes versicolor	ZFA	150	32
1077	gi1621501	Homo sapiens	Human REST protein mRNA, complete cds.	145	32
1077	AAR99365	Homo sapiens	Human REST protein.	145	32
1078	AAY59682	Homo sapiens	Secreted protein 108-009-5-0-A2-FL.	951	99
1078	AAY01635	Homo sapiens	Human PS214 derived polypeptide.	951	99
1078	AAY64650	Homo sapiens	Human human homology protein.	951	99
1079	gi12002058	Homo sapiens	p5326 mRNA, complete cds.	1339	100
1079	AAB36595	Homo sapiens	Human FLEXHT-17 protein sequence SEQ ID NO:17.	1339	100
1079	gi13325192	Homo sapiens	clone IMAGE:3640823, mRNA, partial cds.	1302	100
1080	gi14041646	Homo sapiens	mRNA for OTT-MAL protein resulting from variant (1;22) translocation breakpoint.	468	42
1080	gi14161367	Homo sapiens	megakaryoblastic leukemia-1 protein (MKL1) mRNA, complete cds.	468	42
1080	gi14041618	Homo sapiens	mRNA for megacaryocytic acute leukemia (MAL gene).	468	42
1081	gi7020770	Homo sapiens	cDNA FLJ20571 fis, clone REC01040.	746	88
1081	gi12653983	Homo sapiens	HSPC142 protein, clone MGC:774 IMAGE:3504488, mRNA, complete cds.	746	88
1081	gi13623287	Homo sapiens	HSPC142 protein, clone MGC:11295 IMAGE:3947341,	746	88

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			mRNA, complete cds.		
1082	AAB75321	Homo sapiens	Human secreted protein sequence encoded by gene 34 SEQ ID NO:140.	507	100
1082	gi3875721	Caenorhabditis elegans	F11C1.1	155	32
1082	gi817987	Mus musculus	neural cell adhesion molecule NCAM-140 and NCAM-180	77	41
1083	gi10433795	Homo sapiens	cDNA FLJ12343 fis, clone MAMMA1002292.	2919	100
1083	AAB93956	Homo sapiens	Human protein sequence SEQ ID NO:14000.	2919	100
1083	AAY86552	Homo sapiens	Human gene 79-encoded protein fragment, SEQ ID NO:469.	807	87
1084	gi6855513	Gallus gallus	syndesmos	552	63
1084	gi13623247	Homo sapiens	Similar to RIKEN cDNA 1110001K21 gene, clone MGC:11275 IMAGE:3944355, mRNA, complete cds.	533	61
1084	gi15157405	Agrobacterium tumefaciens	AGR_C_4093p	86	36
1085	gi10436269	Homo sapiens	cDNA FLJ13956 fis, clone Y79AA1001185.	672	100
1085	AAB94844	Homo sapiens	Human protein sequence SEQ ID NO:16020.	672	100
1085	AAG01524	Homo sapiens	Human secreted protein, SEQ ID NO: 5605.	72	36
1086	gi12659138	Mus musculus	mage-d3	149	31
1086	gi13123516	Mus musculus	magphinin-gamma	141	30
1086	gi13123514	Mus musculus	magphinin-beta 2	141	30
1087	gi2695659	Bos taurus	pyruvate dehydrogenase phosphatase regulatory subunit precursor; PDPr	946	80
1087	gi14022085	Mesorhizobium loti	sarcosine dehydrogenase	347	37
1087	gi15075297	Sinorhizobium meliloti	PUTATIVE OXIDOREDUCTASE PROTEIN	344	39
1088	gi10440476	Homo sapiens	mRNA for FLJ00075 protein, partial cds.	843	98
1088	AAY35936	Homo sapiens	Extended human secreted protein sequence, SEQ ID NO. 185.	395	100
1088	AAB64417	Homo sapiens	Amino acid sequence of human intracellular signalling molecule INTRA49.	360	100
1089	gi538413	Mus musculus	zinc finger protein	3113	94
1089	gi186774	Homo sapiens	Human Kruppel related zinc finger protein (HTF10) mRNA, complete cds.	2466	57
1089	gi6088100	Homo sapiens	mRNA for zinc finger protein (ZFD25), complete cds.	2096	52
1090	gi14719305	Mus musculus	SNX20	1242	77
1090	gi14719307	Homo sapiens	SNX21 (SNX21) mRNA, complete cds.	404	37
1090	gi14149068	Homo sapiens	Novel human gene mapping to chromosome 20.	404	37

143
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1091	gi14597912	Homo sapiens	human CLASP-3	6776	99
1091	gi14598037	Homo sapiens	human CLASP-7	4298	71
1091	gi10440381	Homo sapiens	mRNA for FLJ00026 protein, partial cds.	3676	60
1092	gi2425111	Dictyostelium discoideum	ZipA	168	21
1092	gi167835	Dictyostelium discoideum	myosin heavy chain	105	20
1092	gi3044185	Plasmodium falciparum	mature parasite-infected erythrocyte surface antigen	127	22
1093	AAB63245	Homo sapiens	Human breast cancer associated antigen protein sequence SEQ ID NO:607.	1775	100
1093	gi15079591	Homo sapiens	Similar to cyclin D3, clone MGC:10400 IMAGE:3946312, mRNA, complete cds.	1514	100
1093	gi181247	Homo sapiens	cyclin D3 (CCND3) mRNA, complete cds.	1514	100
1094	gi8926320	Rattus norvegicus	corneal wound healing related protein	3609	95
1094	gi10437745	Homo sapiens	cDNA: FLJ21613 fis, clone COL07381.	2655	98
1094	gi10439113	Homo sapiens	cDNA: FLJ22643 fis, clone HSI07031.	1116	100
1095	gi7023247	Homo sapiens	cDNA FLJ10908 fis, clone OVARC1000087, weakly similar to HISTONE MACRO-H2A.1.	2776	99
1095	AAB93319	Homo sapiens	Human protein sequence SEQ ID NO:12406.	2776	99
1095	gi10433992	Homo sapiens	cDNA FLJ12480 fis, clone NT2RM1001066.	2227	100
1096	AAB43434	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:879.	545	83
1096	AAB57205	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1783.	459	81
1096	gi13097207	Homo sapiens	ribosomal protein, large, P1, clone MGC:5215 IMAGE:2900846, mRNA, complete cds.	452	84
1097	gi11138036	Homo sapiens	mRNA, similar to rat myomegalin, complete cds.	885	100
1097	gi13543594	Homo sapiens	similar to rat myomegalin, clone MGC:14586 IMAGE:4247433, mRNA, complete cds.	865	98
1097	gi11138042	Homo sapiens	mRNA, similar to rat myomegalin, complete cds.	833	97
1098	gi7332088	Caenorhabditis elegans	contains similarity to Pfam family PF01391 (Collagen triple helix repeat (20 copies)), score=73.8, E=3.5e-18, N=2	98	36
1098	gi4218141	Antirrhinum majus	HMR1 protein	96	34
1098	gi156262	Caenorhabditis elegans	collagen	93	35
1099	gi12224960	Homo sapiens	mRNA; cDNA DKFZp547J036 (from clone DKFZp547J036).	1336	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1099	gi15145797	Sus scrofa	basic proline-rich protein	143	32
1099	gi904359	Beta vulgaris	chitinase 1	137	39
1100	gi8163691	Streptococcus pneumoniae	surface protein PspC	110	26
1100	gi433475	Drosophila melanogaster	protein 34-specific exons	97	24
1100	gi158694	Drosophila melanogaster	tropomyosin isoform 34 (9B)	86	34
1101	gi467067	Mycobacterium leprae	B2168_F1_37	100	27
1101	gi1296947	Tentyria schaumii	cytochrome oxidase I	68	42
1101	gi1296883	Hegeter glaber	cytochrome oxidase I	63	40
1102	gi10435272	Homo sapiens	cDNA FLJ13305 fis, clone OVARC1001399.	2297	99
1102	AAB94564	Homo sapiens	Human protein sequence SEQ ID NO:15341.	2297	99
1102	gi2947228	Plasmodium yoelii yoelii	erythrocyte binding protein	164	19
1103	gi13276571	Homo sapiens	partial mRNA for chr10 synaptotagmin (CHR10SYT gene).	824	98
1103	gi1932801	Rattus norvegicus	synaptotagmin X	381	35
1103	gi6136792	Mus musculus	synaptotagmin X	379	35
1104	AAY73389	Homo sapiens	HTRM clone 4173111 protein sequence.	1023	100
1104	AAB32463	Homo sapiens	Human secreted protein sequence encoded by gene 22 SEQ ID NO:149.	774	100
1104	AAB32465	Homo sapiens	Human secreted protein sequence encoded by gene 22 SEQ ID NO:151.	235	100
1105	gi13543642	Homo sapiens	clone MGC:14625 IMAGE:4076810, mRNA, complete cds.	561	89
1105	gi2343185	Homo sapiens	tubulin folding cofactor B mRNA, complete cds.	560	90
1105	AAW47209	Homo sapiens	Homo sapiens tubulin-folding cofactor B.	560	90
1106	gi3790545	Mus musculus	neuronal protein 4.1	3665	95
1106	gi4587118	Rattus norvegicus	rat brain 4.1(S)	3660	91
1106	gi4587120	Rattus norvegicus	rat brain 4.1(L)	2882	79
1107	gi3790545	Mus musculus	neuronal protein 4.1	4367	95
1107	gi4587118	Rattus norvegicus	rat brain 4.1(S)	4361	95
1107	gi4587120	Rattus norvegicus	rat brain 4.1(L)	2882	79
1108	gi5912239	Homo sapiens	mRNA; cDNA DKFZp434O225 (from clone DKFZp434O225); partial cds.	290	34
1108	gi5912034	Homo sapiens	mRNA; cDNA DKFZp434N0535 (from clone DKFZp434N0535); partial cds.	290	34
1108	gi7023228	Homo sapiens	cDNA FLJ10898 fis, clone	230	48

145
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			NT2RP5003492.		
1109	gi10179839	Homo sapiens	Thy-1 glycoprotein and Thy-1 co-transcribed protein mRNAs, complete cds.	761	100
1109	gi2944066	Canis familiaris	synapsin I	110	33
1109	gi530876	Chlamydomonas reinhardtii	amino acid feature: Rod protein domain, aa 266 .. 468; amino acid feature: globular protein domain, aa 32 .. 265	99	29
1110	gi15144271	Homo sapiens	NYD-SP11 mRNA, complete cds.	2052	61
1110	gi10178281	Arabidopsis thaliana	katanin p80 subunit-like protein	130	25
1110	gi14578567	Schizosaccharomyces pombe	Shk1 kinase binding protein 15	109	20
1111	gi15144273	Homo sapiens	protein kinase NYD-SP24 mRNA, complete cds.	1563	100
1111	gi10437702	Homo sapiens	cDNA: FLJ21579 fis, clone COL06761.	1415	100
1111	gi14647541	Zapus hudsonius	growth hormone receptor	80	37
1112	AAB95856	Homo sapiens	Human protein sequence SEQ ID NO:18916.	109	28
1112	gi3885945	Ornithorhynchus anatinus	amelogenin	94	28
1112	gi3885947	Tachyglossus aculeatus	amelogenin	94	29
1113	gi10433312	Homo sapiens	cDNA FLJ11939 fis, clone HEMBB1000592.	1110	100
1113	AAB93872	Homo sapiens	Human protein sequence SEQ ID NO:13771.	1110	100
1113	gi560496	Cercopithecine herpesvirus 1	glycoprotein G (homologue of HSV-2 US4)	109	34
1114	gi10433986	Homo sapiens	cDNA FLJ12476 fis, clone NT2RM1000978.	3253	100
1114	AAB94049	Homo sapiens	Human protein sequence SEQ ID NO:14213.	3253	100
1114	gi15080775	Homo sapiens	protein kinase NYD-SP5 mRNA, complete cds.	1565	98
1115	gi4159888	Homo sapiens	PAC clone RP5-855D21, complete sequence.	2212	99
1115	gi1769491	Homo sapiens	Human kruppel-related zinc finger protein (ZNF184) mRNA, partial cds.	1127	49
1115	gi13752754	Homo sapiens	zinc finger 1111 mRNA, complete cds.	1077	47
1116	gi11464740	Homo sapiens	gigaxonin mRNA, complete cds.	3160	100
1116	gi3789797	Homo sapiens	actin binding protein MAYVEN mRNA, complete cds.	633	28
1116	gi6644176	Homo sapiens	kelch-like protein KLHL3a (KLHL3a) mRNA, complete cds.	593	28
1117	gi13278963	Homo sapiens	clone IMAGE:3603836, mRNA, partial cds.	1224	99
1117	AAG02054	Homo sapiens	Human secreted protein, SEQ ID NO: 6135.	678	95
1117	gi9757771	Arabidopsis thaliana	RING zinc finger protein-like	318	34

146
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1118	gi3869259	Homo sapiens	ZNF202 beta (ZNF202) mRNA, complete cds.	684	40
1118	gi7328045	Homo sapiens	mRNA; cDNA DKFZp727E211 (from clone DKFZp727E211); partial cds.	684	40
1118	AAY29917	Homo sapiens	Human CHD1 protein encoded by transcript cDNA1.	684	40
1119	AAV48125_aal	Homo sapiens	Nucleotide sequence encoding clone HMWGS46 of Prohibitin receptor family.	1297	89
1119	gi2289906	Mus musculus	BAP	1296	89
1119	gi6563274	Homo sapiens	B-cell receptor-associated protein BAP37 mRNA, complete cds.	1296	89
1120	gi13540403	Lens culinaris	histone H1	104	28
1120	gi13425157	Caulobacter crescentus	arylesterase-related protein	99	28
1120	gi7331176	Homo sapiens	BM88 antigen mRNA, complete cds.	89	31
1121	AAB58852	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 560.	511	100
1121	AAY94955	Homo sapiens	Human secreted protein clone kh13_4 protein sequence SEQ ID NO:116.	66	34
1121	AAW73411	Homo sapiens	Human secreted protein encoded by Gene No. 15.	61	29
1122	gi13938261	Homo sapiens	clone MGC:15514 IMAGE:3028040, mRNA, complete cds.	1370	59
1122	gi5262557	Homo sapiens	mRNA; cDNA DKFZp569D2231 (from clone DKFZp569D2231); partial cds.	1370	59
1122	gi5441615	Canis familiaris	zinc finger protein	1365	56
1123	AAG02740	Homo sapiens	Human secreted protein, SEQ ID NO: 6821.	458	98
1123	AAB65850	Homo sapiens	Human TANGO 281 extracellular domain SEQ ID NO: 51.	70	31
1123	gi1185157	Odontella sinensis	50S ribosomal protein L29	64	22
1124	gi2576348	Homo sapiens	Human Chromosome 16 BAC clone CIT987SK-A-735G6, complete sequence.	1520	100
1124	gi13775464	Caenorhabditis elegans	Contains similarity to Pfam domain: PF00749 (tRNA-synt_1c), Score=323.3, E-value=8.9e-94, N=1	806	38
1124	gi2688265	Borrelia burgdorferi	glutamyl-tRNA synthetase (gltX)	765	36
1125	gi12654783	Homo sapiens	Similar to loss of heterozygosity, 11, chromosomal region 2, gene A, clone MGC:4904 IMAGE:3461486, mRNA, complete cds.	2141	100
1125	AAB82047	Homo sapiens	Human mast cell surface antigen.	2141	100
1125	gi2190974	Homo sapiens	breast cancer suppressor candidate 1 (bcsc-1) mRNA, complete cds.	338	100
1126	gi4884114	Homo sapiens	mRNA; cDNA DKFZp586G0518	4805	99

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			(from clone DKFZp586G0518); partial cds.		
1126	AAB95305	Homo sapiens	Human protein sequence SEQ ID NO:17544.	3327	99
1126	AAB92704	Homo sapiens	Human protein sequence SEQ ID NO:11106.	2993	99
1127	AAAY25717	Homo sapiens	Human secreted protein encoded from gene 7.	66	29
1127	gi7706791	Arabidopsis thaliana	3-hydroxy-3-methylglutaryl coenzyme A reductase isoform HMGR1L	63	31
1127	gi14861050	Ornithorhynchus anatinus	sulfotransferase SULT1A	76	39
1128	gi9501358	Mus musculus	Pax1 protein	1683	90
1128	gi12658965	Mus musculus	PAX1	1683	90
1128	gi200224	Mus musculus	Pax-1 protein	1665	90
1129	AAW82571	Homo sapiens	Human BBP1 DNA.	1829	93
1129	AAW82572	Homo sapiens	Human BBP1 protein fragment.	417	97
1129	gi563377	Homo sapiens	H.sapiens (JER58) MUC5AC mRNA for mucin (partial).	199	25
1130	AAAY10834	Homo sapiens	Amino acid sequence of a human secreted protein.	1603	99
1130	gi10433027	Homo sapiens	cDNA FLJ11712 fis, clone HEMBA1005185.	1597	99
1130	AAB93801	Homo sapiens	Human protein sequence SEQ ID NO:13573.	1597	99
1131	AAAY10834	Homo sapiens	Amino acid sequence of a human secreted protein.	1565	100
1131	gi10433027	Homo sapiens	cDNA FLJ11712 fis, clone HEMBA1005185.	1559	99
1131	AAB93801	Homo sapiens	Human protein sequence SEQ ID NO:13573.	1559	99
1132	gi13195147	Mus musculus	HCH	9033	94
1132	gi1339910	Homo sapiens	Human DOCK180 protein mRNA, complete cds.	5566	61
1132	AAW03515	Homo sapiens	Human DOCK180 protein.	5563	61
1133	gi12053209	Homo sapiens	mRNA; cDNA DKFZp434E2135 (from clone DKFZp434E2135); complete cds.	549	100
1133	gi695686	Picea abies	dall	59	42
1133	gi4894319	Bacillus anthracis	pXO1-104	53	30
1134	gi3411250	Dictyostelium discoideum	developmental protein	94	36
1134	gi3600087	Dictyostelium discoideum	cytosolic regulator pianissimo	94	36
1135	gi13543686	Homo sapiens	Similar to RIKEN cDNA 4931428F02 gene, clone MGC:14797 IMAGE:4064169, mRNA, complete cds.	1486	98
1135	AAAY07081	Homo sapiens	Renal cancer associated antigen precursor sequence.	261	40
1135	gi7020359	Homo sapiens	cDNA FLJ20333 fis, clone HEP11252.	206	31
1136	gi14250138	Homo sapiens	Similar to RIKEN cDNA	1377	100

148
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			5730421E18 gene, clone MGC:14798 IMAGE:4064321, mRNA, complete cds.		
1136	gi1256896	Saccharomyces cerevisiae	Ylr143wp	552	49
1136	gi995714	Saccharomyces cerevisiae	L3177	552	49
1137	gi10798580	Homo sapiens	mRNA for 17kD fetal brain protein.	410	93
1137	AAY57916	Homo sapiens	Human transmembrane protein HTMPN-40.	410	93
1137	gi1498731	Brassica napus	pathogenesis-related protein PR1	111	35
1138	gi6934248	Homo sapiens	tropomodulin 4 (TMOD4) mRNA, complete cds.	1081	83
1138	gi12744762	Homo sapiens	muscle tropomodulin (TMOD4) gene, complete cds.	1081	83
1138	gi6013189	Homo sapiens	Sk-tropomodulin (Sk-Tmod) mRNA, complete cds.	1081	83
1139	gi1841302	Sus scrofa	ribosomal protein	516	89
1139	gi206732	Rattus norvegicus	ribosomal protein L36a	516	89
1139	gi1666702	Mus musculus	ribosomal protein	516	89
1140	gi12653007	Homo sapiens	clone MGC:3180 IMAGE:3356277, mRNA, complete cds.	1220	100
1140	AAG00664	Homo sapiens	Human secreted protein, SEQ ID NO: 4745.	619	100
1140	AAG00663	Homo sapiens	Human secreted protein, SEQ ID NO: 4744.	489	100
1141	gi14250752	Homo sapiens	Similar to hect domain and RLD 2, clone MGC:14386 IMAGE:4300738, mRNA, complete cds.	890	100
1141	gi4079809	Homo sapiens	HERC2 (HERC2) mRNA, complete cds.	264	28
1141	gi5478530	Arabidopsis thaliana	UVB-resistance protein UVR8	270	32
1142	gi11322969	Homo sapiens	HCG V gene.	485	90
1142	gi11322968	Homo sapiens	HCG V mRNA.	485	90
1142	gi3176438	Homo sapiens	MHC class 1 region.	485	90
1143	gi12224835	Homo sapiens	mRNA; cDNA DKFZp547I082 (from clone DKFZp547I082).	622	100
1143	AAB51649	Homo sapiens	Human secreted protein sequence encoded by gene 30 SEQ ID NO:89.	272	100
1143	AAB75511	Homo sapiens	Human secreted protein sequence encoded by gene 6 SEQ ID NO:65.	272	100
1144	AAB43827	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1272.	509	87
1144	gi13775484	Caenorhabditis elegans	similar to plasmodium merozoite surface antigen precursor (SP:P04933)	359	39
1144	gi6808143	Homo sapiens	mRNA; cDNA DKFZp761D051 (from clone DKFZp761D051); partial cds.	162	36
1145	gi14549207	Homo sapiens	URAX1 mRNA, complete cds.	3118	100
1145	AAB27591	Homo sapiens	Human secreted protein SEQ ID NO: 92.	2466	98

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1145	gi5901836	Drosophila melanogaster	BcDNA.GH09817	403	49
1146	gi1492084	Molluscum contagiosum virus subtype 1	MC141R	66	46
1146	AAY72171	Homo sapiens	Human RNA metabolism protein (RMEP-11).	58	36
1146	gi12722186	Pasteurella multocida	SelA	66	31
1147	AAB53429	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:969.	1118	98
1147	AAG02922	Homo sapiens	Human secreted protein, SEQ ID NO: 7003.	60	43
1147	gi3123743	synthetic construct	echistatin	58	36
1149	gi52977	Mus musculus	modifier 3 (M33)	2187	83
1149	gi2266988	Mus musculus	M33 polycomb-like protein	2167	82
1149	gi3860185	Xenopus laevis	Polycomb homolog Pc1	748	49
1150	gi15213542	Homo sapiens	NSD1 (NSD1) mRNA, complete cds.	1618	100
1150	gi3329465	Mus musculus	NSD1 protein	1458	90
1150	gi3249715	Homo sapiens	MMSET type I (WHSC1) mRNA, complete cds.	86	31
1151	gi12005728	Homo sapiens	GL012 mRNA, complete cds.	922	100
1151	gi258834	Mus sp.	alpha 1 (XII) collagen {triple-helical domain COL2}	61	37
1151	gi4583418	Homo sapiens	Kruppel-like zinc finger transcription factor gene, complete cds.	77	46
1152	AAB47134	Homo sapiens	CDIFF-15, Incyte ID No. 3478571CD1.	1286	59
1152	AAY27621	Homo sapiens	Human secreted protein encoded by gene No. 55.	661	70
1152	gi7022185	Homo sapiens	cDNA FLJ10260 fis, clone HEMBB1000973, moderately similar to Mus musculus schlafen3 mRNA.	379	44
1153	gi12654519	Homo sapiens	clone IMAGE:3504989, mRNA, partial cds.	1809	99
1153	gi10440173	Homo sapiens	cDNA: FLJ23471 fis, clone HSI11969.	470	29
1153	gi10440416	Homo sapiens	mRNA for FLJ00043 protein, partial cds.	257	45
1154	AAG03284	Homo sapiens	Human secreted protein, SEQ ID NO: 7365.	364	100
1154	gi2853301	Homo sapiens	mucin (MUC3) mRNA, partial cds.	111	21
1154	gi2586071	Xenopus laevis	kinesin-related protein	95	25
1155	AAB58405	Homo sapiens	Lung cancer associated polypeptide sequence SEQ ID 743.	814	99
1155	gi9105937	Xylella fastidiosa 9a5c	acetylglutamate kinase	505	36
1155	gi5155	Saccharomyces cerevisiae	acetylglutamate kinase	312	28
1156	gi13436152	Homo sapiens	reticulocalbin 2, EF-hand calcium binding domain, clone MGC:1650	1678	100

150
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			IMAGE:3505241, mRNA, complete cds.		
1156	gi469885	Homo sapiens	H.sapiens ERC-55 mRNA.	1678	100
1156	AAW21949	Homo sapiens	E6-binding protein E6-BPSD7.	1678	100
1157	gi1293893	Mus musculus	leucine zipper protein 1	4333	80
1157	gi14042584	Homo sapiens	cDNA FLJ14799 fis, clone NT2RP4001351, weakly similar to Human ovarian cancer downregulated myosin heavy chain homolog (Doc1) mRNA.	604	37
1157	AAB93244	Homo sapiens	Human protein sequence SEQ ID NO:12253.	604	37
1158	gi9964287	Homo sapiens	hypertension-related calcium-regulated gene mRNA, complete cds.	1036	94
1158	gi10434820	Homo sapiens	cDNA FLJ13008 fis, clone NT2RP3000456.	1036	94
1158	gi12803673	Homo sapiens	HT002 protein; hypertension-related calcium-regulated gene, clone MGC:3418 IMAGE:3606279, mRNA, complete cds.	1036	94
1159	gi3387790	Homo sapiens	protein tyrosine phosphatase PIR1 mRNA, complete cds.	1135	97
1159	gi12653157	Homo sapiens	dual specificity phosphatase 11 (RNA/RNP complex 1-interacting), clone MGC:8576 IMAGE:2960939, mRNA, complete cds.	1127	97
1159	AAG01376	Homo sapiens	Human secreted protein, SEQ ID NO: 5457.	542	100
1160	gi5901688	Mus musculus	GRIN1	217	22
1160	gi557822	Saccharomyces cerevisiae	mal5, sta1, len: 1367, CAI: 0.3, AMYH_YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	206	20
1160	gi1304387	Saccharomyces cerevisiae var. diastaticus	glucoamylase	206	20
1161	gi12803603	Homo sapiens	clone MGC:4161 IMAGE:3606994, mRNA, complete cds.	2643	100
1161	AAB21042	Homo sapiens	Human nucleic acid-binding protein, NuABP-46.	397	51
1161	gi4097497	Mus musculus	zinc finger protein 94	378	39
1162	gi13661556	Homo sapiens	extracellular glycoprotein EMILIN-2 precursor, mRNA, complete cds.	1410	98
1162	gi14517602	Homo sapiens	mRNA for FOAP-10 protein, partial cds.	1092	99
1162	gi14043093	Homo sapiens	clone MGC:15203 IMAGE:3163767, mRNA, complete cds.	362	45
1163	gi311946	Homo sapiens	H.sapiens p18 mRNA.	85	31
1163	gi4927776	Homo sapiens	alternative HHLA3 protein (HHLA3) mRNA, complete cds.	63	36
1163	gi13883114	Mycobacterium tuberculosis CDC1551	helicase, UvrD/Rep family	83	43
1164	gi12053149	Homo sapiens	mRNA; cDNA DKFZp434G2226	4649	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			(from clone DKFZp434G2226); complete cds.		
1164	gi2239242	Schizosaccharomyces pombe	kinesin-like protein	954	42
1164	gi1881662	Drosophila melanogaster	kinesin like protein 67a	962	43
1165	gi15186738	Mus musculus	Tac2-N	2229	87
1165	AAG03542	Homo sapiens	Human secreted protein, SEQ ID NO: 7623.	408	97
1165	gi6136786	Mus musculus	synaptotagmin VII	207	25
1166	gi2760351	Girardia tigrina	myosin heavy chain	197	23
1166	gi3986194	Dugesia japonica	myosin heavy chain	191	22
1166	gi11067	Schistosoma mansoni	myosin II heavy chain	182	21
1167	AAY56021	Homo sapiens	Human CD40 receptor interacting protein 4C4.	2121	99
1167	gi3638955	Homo sapiens	PAC clone RP4-751H13 from 7q35-qter, complete sequence.	298	22
1167	gi4063770	Ipomoea purpurea	transposase	130	23
1168	gi7022811	Homo sapiens	cDNA FLJ10649 fis, clone NT2RP2005835, weakly similar to SHP1 PROTEIN.	116	26
1168	AAB93031	Homo sapiens	Human protein sequence SEQ ID NO:11803.	116	26
1168	gi7800642	Streptococcus pneumoniae	PspA	121	39
1169	gi15081398	Homo sapiens	kruppel-like zinc finger protein (ZNF300) mRNA, complete cds.	218	53
1169	gi340448	Homo sapiens	Human zinc-finger protein 8 (ZFP8) mRNA, 3' end.	211	64
1169	gi14602511	Homo sapiens	Similar to zinc finger protein 274, clone MGC:1159 IMAGE:2966764, mRNA, complete cds.	201	52
1170	gi22333	Zea mays	hydroxyproline-rich glycoprotein	90	24
1170	gi4007865	Zea mays	Hydroxyproline-rich Glycoprotein (HRGP)	90	26
1170	gi257041	Zea mays	hydroxyproline-rich glycoprotein, HRGP	90	26
1171	gi10440267	Homo sapiens	cDNA: FLJ23544 fis, clone LNG08336.	695	100
1171	AAG02534	Homo sapiens	Human secreted protein, SEQ ID NO: 6615.	61	40
1172	gi13436440	Homo sapiens	clone MGC:4400 IMAGE:2905976, mRNA, complete cds.	2543	99
1172	gi10434258	Homo sapiens	cDNA FLJ12644 fis, clone NT2RM4001979, weakly similar to ZINC FINGER PROTEIN 84.	909	57
1172	AAB94205	Homo sapiens	Human protein sequence SEQ ID NO:14550.	909	57
1173	gi6855513	Gallus gallus	syndesmos	417	64
1173	gi13623247	Homo sapiens	Similar to RIKEN cDNA 1110001K21 gene, clone MGC:11275 IMAGE:3944355, mRNA, complete cds.	403	62

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1173	gi15157405	Agrobacterium tumefaciens	AGR_C_4093p	86	36
1174	gi13310782	Mus musculus	myoneurin	237	33
1174	gi498736	Homo sapiens	H.sapiens HZF9 mRNA for zinc finger protein.	227	51
1174	gi10086309	Mus musculus	Kruppel-type zinc finger protein KROX-25	227	57
1175	AAB58899	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 607.	843	99
1175	gi9967699	Schizosaccharom yces pombe	protein with weak similarity to Azospirillum brasilense nifR3 protein; yeast YML080W homolog; UPF0034 domain	732	50
1175	gi15146316	Arabidopsis thaliana	AT5g67220/K21H1_18	687	51
1176	gi11558109	Homo sapiens	mRNA for PRAM-1 protein.	2373	87
1176	gi13938357	Homo sapiens	Similar to RIKEN cDNA 0610030H11 gene, clone MGC:15716 IMAGE:3353224, mRNA, complete cds.	943	67
1176	gi5917666	Zea mays	extensin-like protein	466	28
1177	gi10432603	Homo sapiens	cDNA FLJ11360 fis, clone HEMBA1000231.	3044	100
1177	AAB93711	Homo sapiens	Human protein sequence SEQ ID NO:13305.	3044	100
1177	gi13872809	Homo sapiens	mRNA for artemis.	3001	99
1178	gi12382773	Homo sapiens	caspase recruitment domain protein 11 mRNA, complete cds.	1999	99
1178	gi13488607	Homo sapiens	caspase recruitment domain protein 10 mRNA, complete cds.	309	33
1178	gi14192725	Homo sapiens	CARD-containing MAGUK 3 protein (CARMA3) mRNA, complete cds.	309	33
1179	gi12652571	Homo sapiens	integral type I protein, clone MGC:1302 IMAGE:3507308, mRNA, complete cds.	691	100
1179	gi14790009	Homo sapiens	Similar to integral type I protein, clone MGC:8938 IMAGE:3876465, mRNA, complete cds.	691	100
1179	gi4583677	Homo sapiens	mRNA for integral type I protein p24B (p26).	691	100
1180	AAB47327	Homo sapiens	FCTR4.	3068	99
1180	gi3041860	Homo sapiens	PAC clone RP4-545C24 from 7q21-q22, complete sequence.	1275	53
1180	gi14603161	Homo sapiens	Rho guanine nucleotide exchange factor (GEF) 5, clone MGC:19765 IMAGE:3636857, mRNA, complete cds.	1275	53
1181	AAW79092	Homo sapiens	Human secreted protein dn740_3.	724	70
1181	AAV73380	Homo sapiens	HTRM clone 1821233 protein sequence.	724	70
1181	gi13543342	Homo sapiens	clone IMAGE:2821841, mRNA, partial cds.	585	36
1182	gi13097648	Homo sapiens	clone IMAGE:3610995, mRNA,	782	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			partial cds.		
1182	gi7019947	Homo sapiens	cDNA FLJ20080 fis, clone COL03184.	112	26
1182	gi10433190	Homo sapiens	cDNA FLJ11837 fis, clone HEMBA1006612.	105	26
1183	gi13097648	Homo sapiens	clone IMAGE:3610995, mRNA, partial cds.	782	100
1183	gi8977890	Homo sapiens	mRNA full length insert cDNA clone EUROIMAGE 364773.	118	30
1183	gi5817164	Homo sapiens	mRNA; cDNA DKFZp434H204 (from clone DKFZp434H204).	118	30
1184	gi10434709	Homo sapiens	cDNA FLJ12934 fis, clone NT2RP2004978, weakly similar to ACTIN-LIKE PROTEIN ARP8.	3306	99
1184	AAB94379	Homo sapiens	Human protein sequence SEQ ID NO:14928.	3306	99
1184	gi10439016	Homo sapiens	cDNA: FLJ22579 fis, clone HSI02562.	1978	100
1185	gi1160355	Caenorhabditis elegans	UNC-89	307	27
1185	gi8250181	Drosophila melanogaster	D-Titin	319	22
1186	gi13276231	Homo sapiens	mRNA for FYVE and coiled-coil domain containing 1 (FYCO1 gene).	7470	100
1186	gi13938195	Homo sapiens	FYVE and coiled-coil domain containing 1, clone MGC:14998 IMAGE:3050704, mRNA, complete cds.	1271	98
1186	gi10435317	Homo sapiens	cDNA FLJ13335 fis, clone OVARC1001861.	1263	99
1187	gi10435317	Homo sapiens	cDNA FLJ13335 fis, clone OVARC1001861.	1290	100
1187	AAB94590	Homo sapiens	Human protein sequence SEQ ID NO:15398.	1290	100
1187	gi13938195	Homo sapiens	FYVE and coiled-coil domain containing 1, clone MGC:14998 IMAGE:3050704, mRNA, complete cds.	1265	98
1188	gi1235527	Homo sapiens	Human myoglobin gene (exon 1) (and joined CDS).	809	100
1188	gi386872	Homo sapiens	Human myoglobin gene, exon 3.	806	99
1188	gi164547	Sus scrofa	myoglobin	758	93
1189	gi14587078	Danio rerio	Unc119c	857	69
1189	gi14587076	Danio rerio	Unc119b	782	73
1189	gi1161378	Rattus norvegicus	retinal protein	700	71
1190	gi404751	Mus musculus	adenylosuccinate synthetase	2353	96
1190	gi415849	Homo sapiens	Human adenylosuccinate synthetase mRNA.	1864	77
1190	gi404057	Mus musculus	adenylosuccinate synthetase	1842	76
1191	gi15081791	Arabidopsis thaliana	At1g09280/T12M4_1	527	36
1191	gi12541446	Corynebacterium glutamicum	RXC00552	447	34
1191	gi13702647	Staphylococcus	conserved hypotehtical protein	439	34

154
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		aureus subsp. aureus N315			
1192	gi2970646	Mus musculus	Xin	3622	53
1192	gi2970644	Gallus gallus	Xin	1513	47
1192	gi15029903	Mus musculus	Similar to proline-rich protein BstNI subfamily 2	202	27
1193	gi12804921	Homo sapiens	Similar to membrane protein of cholinergic synaptic vesicles, clone MGC:2671 IMAGE:3547155, mRNA, complete cds.	1556	100
1193	gi14250546	Homo sapiens	clone MGC:3015 IMAGE:3162543, mRNA, complete cds.	1556	100
1193	gi1698401	Homo sapiens	Human BRCA1, Rho7 and vatI genes, complete cds, and ipf35 gene, partial cds.	1553	99
1194	gi14598037	Homo sapiens	human CLASP-7	4269	99
1194	gi14249911	Homo sapiens	clone IMAGE:3506202, mRNA, partial cds.	3528	100
1194	gi14597918	Homo sapiens	human CLASP-3	3156	72
1195	gi11493417	Homo sapiens	PRO1292	244	100
1195	gi1574381	Haemophilus influenzae Rd	lic-1 operon protein (licC)	39	37
1195	gi13359399	Oserya coulteriana	maturase K	41	53
1196	AAB60467	Homo sapiens	Human cell cycle and proliferation protein CCYPR-15, SEQ ID NO:15.	972	100
1196	AAB75130	Homo sapiens	Human SCAP 22 protein sequence SEQ ID NO:2.	968	99
1196	gi12805505	Mus musculus	Similar to CHMP1.5 protein	958	97
1197	gi11559490	Homo sapiens	VDUP1 gene, complete cds.	1631	99
1197	gi7717464	Homo sapiens	brain-expressed HHCPA78 homolog VDUP1 (Gene) mRNA, complete cds.	1631	99
1197	gi11181624	Mus musculus	thioredoxin interacting factor	1580	95
1198	gi14906268	Homo sapiens	Pur-beta (PURB) mRNA, complete cds.	1638	100
1198	gi2460119	Mus musculus	vascular actin single-stranded DNA-binding factor 2 p44 component; purine-rich single-stranded DNA-binding protein beta; PurB beta	1569	94
1198	gi190750	Homo sapiens	H.sapiens Pur (pur-alpha) mRNA, complete cds.	979	71
1199	gi10438372	Homo sapiens	cDNA: FLJ22104 fis, clone HEP17633.	1867	99
1199	gi5441952	Homo sapiens	peroxisomal membrane protein PMP 24 mRNA, complete cds.	87	28
1199	gi12654621	Homo sapiens	24 kDa intrinsic membrane protein, clone MGC:1213 IMAGE:3533572, mRNA, complete cds.	87	28
1200	gi1039447	Saccharomyces cerevisiae	Lpb1p	243	43
1200	gi1171427	Saccharomyces cerevisiae	Ypl030wp	121	46
1200	gi6318215	Mus musculus	E-selectin ligand 1 (ESL-1)	97	31
1201	gi13529542	Mus musculus	RIKEN cDNA 4731402F03 gene	609	45

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1201	AAW50193	Homo sapiens	Amino acid sequence of salivary protein CON-2.	74	48
1201	gi11559406	Canis familiaris	dopamine receptor D4	86	57
1202	gi13561413	Homo sapiens	F-BOX domain protein mRNA, complete cds.	2076	99
1202	gi6561831	Homo sapiens	muscle disease-related protein mRNA, complete cds.	682	40
1202	gi1651195	Mus musculus	TRAF6	96	38
1203	gi219980	Homo sapiens	Human mRNA for paraneoplastic cerebellar degeneration-associated antigen, complete cds.	2243	100
1203	gi241777	human, mRNA, 2192 nt]. [Homo sapiens	28S RNA, autoantigen recognized by an anti-neuronal cell antibody	2243	100
1203	gi180187	Homo sapiens	Human major Yo paraneoplastic antigen (CDR2) mRNA, 3' end.	2206	96
1204	gi11870006	Rattus norvegicus	nuclear matrix transcription factor	737	81
1204	gi6729087	Rattus norvegicus	Cas-associated zinc finger protein	719	80
1204	gi11870000	Rattus norvegicus	nuclear matrix transcription factor	719	80
1205	gi14043225	Homo sapiens	clone MGC:15678 IMAGE:3350074, mRNA, complete cds.	470	100
1205	gi13537206	Homo sapiens	hMBLR mRNA, complete cds.	467	85
1205	gi14042885	Homo sapiens	cDNA FLJ14979 fis, clone Y79AA1000037, moderately similar to DNA-BINDING PROTEIN BMI-1.	467	85
1206	gi12052744	Homo sapiens	mRNA; cDNA DKFZp564K0322 (from clone DKFZp564K0322); complete cds.	2777	99
1206	gi393194	Plasmodium falciparum	S-antigen	178	27
1206	gi160190	Plasmodium cynomolgi	circumsporozoite antigen	180	35
1207	gi12053315	Homo sapiens	mRNA; cDNA DKFZp586K0717 (from clone DKFZp586K0717); complete cds.	2748	97
1207	AAW74802	Homo sapiens	Human secreted protein encoded by gene 73 clone HSQEL25.	2748	97
1207	AAB56895	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1473.	2748	97
1208	gi153781	Streptococcus sp.	protein H precursor	154	26
1208	gi1247735	Streptococcus pyogenes	protein H	154	26
1208	gi557785	Saccharomyces cerevisiae	tropomyosin	132	28
1209	gi14336699	Homo sapiens	16p13.3 sequence section 2 of 8.	101	30
1209	gi10179425	Homo sapiens	Rb-associated protein mRNA, complete cds.	73	25
1209	AAW88833	Homo sapiens	Polypeptide fragment encoded by gene 70.	66	40
1210	gi10435667	Homo sapiens	cDNA FLJ13611 fis, clone	1817	100

156
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			PLACE1010802.		
1210	AAB95495	Homo sapiens	Human protein sequence SEQ ID NO:18037.	1817	100
1210	gi15080536	Homo sapiens	Similar to RIKEN cDNA 2410002O22 gene, clone MGC:21036 IMAGE:4508655, mRNA, complete cds.	1026	100
1211	gi10435667	Homo sapiens	cDNA FLJ13611 fis, clone PLACE1010802.	1772	98
1211	AAB95495	Homo sapiens	Human protein sequence SEQ ID NO:18037.	1772	98
1211	gi15080536	Homo sapiens	Similar to RIKEN cDNA 2410002O22 gene, clone MGC:21036 IMAGE:4508655, mRNA, complete cds.	981	97
1212	gi1216477	Mus musculus	zinc finger protein 60	138	28
1212	gi6760445	Homo sapiens	Smad- and Olf-interacting zinc finger protein mRNA, partial cds.	142	22
1212	gi4753764	Homo sapiens	OZF gene exon 5 (and complete CDS).	128	24
1213	gi268	Bos taurus	cytochrome c oxidase subunit VIb (AA 1-86)	445	90
1213	gi12654383	Homo sapiens	cytochrome c oxidase subunit VIb, clone MGC:3061 IMAGE:3344701, mRNA, complete cds.	423	83
1213	gi12803321	Homo sapiens	cytochrome c oxidase subunit VIb, clone MGC:1153 IMAGE:3347455, mRNA, complete cds.	423	83
1214	gi13279335	Homo sapiens	pleckstrin homology, Sec7 and coiled/coil domains 2 (cytohesin-2), clone MGC:642 IMAGE:3538580, mRNA, complete cds.	1767	100
1214	gi1575766	Homo sapiens	cytohesin-2 mRNA, complete cds.	1767	100
1214	gi3660538	Mus musculus	cytohesin 2	1756	99
1215	gi12052736	Homo sapiens	mRNA; cDNA DKFZp564H1122 (from clone DKFZp564H1122); complete cds.	1702	100
1215	gi15126751	Homo sapiens	clone MGC:17614 IMAGE:3852558, mRNA, complete cds.	901	98
1215	gi13937992	Homo sapiens	clone MGC:14768 IMAGE:4291902, mRNA, complete cds.	1123	100
1216	AAB63313	Homo sapiens	Human breast cancer associated antigen protein sequence SEQ ID NO:675.	154	31
1216	gi14336722	Homo sapiens	16p13.3 sequence section 3 of 8.	149	32
1216	AAB63312	Homo sapiens	Human breast cancer associated antigen protein sequence SEQ ID NO:674.	132	34
1217	gi8096269	Nicotiana tabacum	KED	162	19
1217	gi12803875	Homo sapiens	Similar to splicing factor, arginine/serine-rich 4, clone MGC:3920 IMAGE:3619538,	160	24

157
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			mRNA, complete cds.		
1217	gi1055054	Caenorhabditis elegans	coded for by C. elegans cDNA yk127b8.5; coded for by C. elegans cDNA yk127b8.3	159	26
1218	AAB56606	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1184.	792	96
1218	AAG01959	Homo sapiens	Human secreted protein, SEQ ID NO: 6040.	567	99
1218	gi9802603	Arabidopsis thaliana	T2E6.7	70	29
1219	gi1373252	Mus musculus	STONE14	1270	82
1219	gi1938245	Rattus norvegicus	PKC-zeta-interacting protein (ZIP)	1248	82
1219	gi13543730	Mus musculus	Similar to sequestosome 1	1271	83
1220	gi10440530	Homo sapiens	mRNA for FLJ00115 protein, partial cds.	4008	99
1220	gi5917730	Homo sapiens	F-box protein Lilina (LILINA) mRNA, complete cds.	2467	99
1220	gi8918522	Homo sapiens	PCCX2 mRNA for protein containing CXXC domain 2, partial cds.	870	72
1221	gi15126546	Homo sapiens	clone IMAGE:3858590, mRNA, partial cds.	1137	99
1221	gi13436296	Homo sapiens	clone IMAGE:3610712, mRNA, partial cds.	828	90
1221	gi14039845	Homo sapiens	testes development-related NYD-SP18 mRNA, complete cds.	136	30
1222	gi6599138	Homo sapiens	mRNA; cDNA DKFZp434I036 (from clone DKFZp434I036); partial cds.	102	24
1222	gi4322304	Homo sapiens	translation initiation factor IF2 mRNA, complete cds.	102	24
1222	gi5002645	Homo sapiens	mRNA for translation initiation factor 2 (IF2 gene).	102	24
1223	gi13276645	Homo sapiens	mRNA; cDNA DKFZp761F0123 (from clone DKFZp761F0123); complete cds.	1999	100
1223	gi14790064	Homo sapiens	clone IMAGE:3878600, mRNA, partial cds.	1595	100
1223	gi3116214	Homo sapiens	mRNA for SH3 binding protein, complete cds.	543	45
1224	gi12053055	Homo sapiens	mRNA; cDNA DKFZp434I1116 (from clone DKFZp434I1116); complete cds.	4734	99
1224	gi10438819	Homo sapiens	cDNA: FLJ22429 fis, clone HRC09084.	1450	97
1224	gi10438616	Homo sapiens	cDNA: FLJ22291 fis, clone HRC04410.	1956	99
1225	AAB21042	Homo sapiens	Human nucleic acid-binding protein, NuABP-46.	2498	99
1225	gi8099348	Homo sapiens	zinc finger protein (ZFP) mRNA, complete cds.	960	55
1225	gi9963806	Homo sapiens	zinc finger protein ZNF287 (ZNF287) mRNA, complete cds.	924	56
1226	gi2217970	Homo sapiens	mRNA for Rab9 effector p40,	1116	81

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			complete cds.		
1226	gi12653463	Homo sapiens	Rab9 effector p40, clone MGC:8459 IMAGE:2821471, mRNA, complete cds.	1113	81
1226	AAB58336	Homo sapiens	Lung cancer associated polypeptide sequence SEQ ID 674.	520	62
1227	gi2217970	Homo sapiens	mRNA for Rab9 effector p40, complete cds.	1115	71
1227	gi12653463	Homo sapiens	Rab9 effector p40, clone MGC:8459 IMAGE:2821471, mRNA, complete cds.	1112	71
1227	AAB58336	Homo sapiens	Lung cancer associated polypeptide sequence SEQ ID 674.	812	64
1228	gi3015538	Homo sapiens	nuclear dual-specificity phosphatase (SBF1) mRNA, partial cds.	2977	62
1228	gi3851594	Drosophila melanogaster	SET domain binding factor	1396	43
1228	gi12224976	Homo sapiens	mRNA; cDNA DKFZp667L246 (from clone DKFZp667L246).	1198	100
1229	gi15081398	Homo sapiens	kruppel-like zinc finger protein (ZNF300) mRNA, complete cds.	796	54
1229	gi13752754	Homo sapiens	zinc finger 1111 mRNA, complete cds.	761	54
1229	gi881564	Homo sapiens	Human zinc finger containing protein ZNF157 (ZNF157) mRNA, complete cds.	931	47
1230	gi2351568	Mus musculus	N-RAP	1530	57
1230	gi806562	Homo sapiens	H.sapiens mRNA for nebulin.	533	31
1230	gi1205990	Homo sapiens	Human nebulin mRNA, partial cds.	453	34
1231	gi14250424	Homo sapiens	clone IMAGE:3869590, mRNA, partial cds.	2252	99
1231	gi13491282	Caenorhabditis elegans	contains similarity to Pfam family PF00515 (TPR Domain), score=51.7, E=1.6e-11, N=3	273	23
1231	gi3599670	Dictyostelium discoideum	TRFA	112	26
1232	gi5911990	Homo sapiens	mRNA; cDNA DKFZp434K1235 (from clone DKFZp434K1235); partial cds.	667	100
1232	gi11095215	Streptomyces rishiriensis	cytochrome P-450	113	21
1232	gi6691467	Triticum aestivum	AHM1	123	29
1233	gi15079828	Homo sapiens	clone MGC:19798 IMAGE:3926284, mRNA, complete cds.	1179	100
1233	gi12224994	Homo sapiens	mRNA; cDNA DKFZp564M0163 (from clone DKFZp564M0163); complete cds.	196	28
1233	AAAY73338	Homo sapiens	HTRM clone 2019742 protein sequence.	194	28
1234	gi12005904	Homo sapiens	AD036 mRNA, complete cds.	751	100
1234	gi2073099	Phycomyces blakesleeanus	chitin synthase	81	35
1235	gi4886503	Homo sapiens	mRNA; cDNA DKFZp564C0469	1401	98

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			(from clone DKFZp564C0469); partial cds.		
1235	gi13938110	Mus musculus	Similar to topoisomerase (DNA) II binding protein	204	28
1235	gi3845613	Homo sapiens	mRNA for DNA topoisomerase II binding protein, complete cds.	204	27
1236	gi9945010	Mus musculus	RING-finger protein MURF	1778	95
1236	gi13160388	Homo sapiens	mRNA for RNF30 gene for ring finger protein 30.	1774	99
1236	AAB21048	Homo sapiens	Human nucleic acid-binding protein, NuABP-52.	1744	95
1237	gi11908000	Homo sapiens	BCL-6 corepressor short isoform (BCOR) mRNA, complete cds; alternatively spliced.	5272	100
1237	gi11907998	Homo sapiens	BCL-6 corepressor (BCOR) mRNA, complete cds; alternatively spliced.	5246	99
1237	gi13592175	Leishmania major	ppg3	203	24
1238	gi12655378	Homo sapiens	partial mRNA for oxodicarboxylate carrier (ODC gene).	1552	100
1238	gi12655649	Rattus norvegicus	mitochondrial oxodicarboxylate carrier	1300	82
1238	gi1050774	Saccharomyces cerevisiae	YOR50-12	491	37
1239	gi12053025	Homo sapiens	mRNA; cDNA DKFZp434P0714 (from clone DKFZp434P0714); complete cds.	3599	99
1239	gi1526421	Mus musculus	KAP3B	97	26
1239	gi3645904	Homo sapiens	Human Smg GDS-associated protein SMAP mRNA, complete cds.	99	25
1240	gi3811111	Homo sapiens	Human I(3)mbt protein homolog mRNA, complete cds.	1445	49
1240	AAY01069	Homo sapiens	Human I(3)mbt protein sequence.	1445	49
1240	gi5817146	Homo sapiens	mRNA; cDNA DKFZp586P1522 (from clone DKFZp586P1522); partial cds.	935	50
1241	gi10437973	Homo sapiens	cDNA: FLJ21801 fis, clone HEP00707.	681	99
1241	AAG00343	Homo sapiens	Human secreted protein, SEQ ID NO: 4424.	277	94
1241	gi12052965	Homo sapiens	mRNA; cDNA DKFZp566M1046 (from clone DKFZp566M1046); complete cds.	198	29
1242	gi15029844	Homo sapiens	clone IMAGE:3451454, mRNA, partial cds.	2315	99
1242	gi14043625	Homo sapiens	clone IMAGE:3139971, mRNA, partial cds.	2303	99
1242	gi13325200	Homo sapiens	clone IMAGE:3659680, mRNA, partial cds.	1906	99
1243	gi14330448	Homo sapiens	mRNA for zinc finger protein RINZF (RINZF gene).	4285	95
1243	gi4557143	Rattus norvegicus	zinc finger protein RIN ZF	1668	89
1243	gi10434431	Homo sapiens	cDNA FLJ12752 fis, clone NT2RP2001174, weakly similar to	1343	100

160
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			GASTRULA ZINC FINGER PROTEIN XLCGF46.1.		
1244	AAG03897	Homo sapiens	Human secreted protein, SEQ ID NO: 7978.	551	91
1244	gi15080686	Lentinula edodes	CDC5	91	26
1244	gi14010354	Homo sapiens	tropomyosin 4-anaplastic lymphoma kinase fusion protein minor isoform mRNA, partial cds.	83	30
1245	AAAY35964	Homo sapiens	Extended human secreted protein sequence, SEQ ID NO. 213.	839	99
1245	gi12718208	Neurospora crassa	related to pseudouridine synthase	261	39
1245	gi9951004	Pseudomonas aeruginosa	tRNA pseudouridine 55 synthase	352	41
1246	gi3025832	Onchocerca volvulus	pyrrolidone-rich antigen	68	41
1246	gi15077042	Bos taurus	survival motor neuron protein	83	32
1246	gi915208	Sus scrofa	gastric mucin	91	24
1247	gi55471	Mus musculus	Zfp-29	1203	52
1247	gi6409345	Homo sapiens	zinc finger protein ZNF180 (ZNF180) mRNA, complete cds.	1199	60
1247	gi13879240	Mus musculus	Similar to zinc finger protein 46	1197	55
1248	AAB56403	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:981.	813	98
1248	gi499204	Drosophila melanogaster	D-E-A-D box protein	407	37
1248	gi7268333	Arabidopsis thaliana	ATP-dependent RNA helicase like protein	387	34
1249	gi11096137	Homo sapiens	MAGEF1 (MAGEF1) mRNA, complete cds.	1592	100
1249	gi14603190	Homo sapiens	MAGEF1 protein, clone MGC:19617 IMAGE:3943384, mRNA, complete cds.	1592	100
1249	AAB60476	Homo sapiens	Human cell cycle and proliferation protein CCYPR-24, SEQ ID NO:24.	1592	100
1250	gi14043262	Homo sapiens	Similar to RIKEN cDNA 1500026B10 gene, clone MGC:15737 IMAGE:3355622, mRNA, complete cds.	1076	100
1250	gi14250512	Homo sapiens	clone MGC:15468 IMAGE:2966921, mRNA, complete cds.	1076	100
1250	gi12006223	Homo sapiens	NPD017 mRNA, complete cds.	464	48
1251	gi12652831	Homo sapiens	clone MGC:5242 IMAGE:2900206, mRNA, complete cds.	657	100
1251	AAAY87070	Homo sapiens	Human secreted protein sequence SEQ ID NO:109.	566	98
1251	AAAY87165	Homo sapiens	Human secreted protein sequence SEQ ID NO:204.	566	98
1252	gi6808293	Homo sapiens	mRNA; cDNA DKFZp434A1010 (from clone DKFZp434A1010); partial cds.	2951	76
1252	gi15145797	Sus scrofa	basic proline-rich protein	340	31
1252	gi600118	Zea mays	extensin-like protein	300	26
1253	gi15139362	Mus musculus	aczonin	6067	96

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1253	gi7493836	Rattus norvegicus	multidomain presynaptic cytomatrix protein Piccolo	6046	95
1253	gi6433844	Gallus gallus	aczonin	5521	87
1254	gi15139362	Mus musculus	aczonin	6001	95
1254	gi7493836	Rattus norvegicus	multidomain presynaptic cytomatrix protein Piccolo	5980	95
1254	gi6433844	Gallus gallus	aczonin	5455	86
1255	gi409389	Rattus norvegicus	dC-stretch binding protein (CSBP)	1793	80
1255	gi241478	human, mRNA, 2302 nt]. [Homo sapiens	heterogeneous nuclear ribonucleoprotein complex K	1793	80
1255	AAY39467	Homo sapiens	Heterogenous ribonuclear protein K.	1793	80
1256	AAW47589	Homo sapiens	T-cell receptor beta-chain.	537	91
1256	gi1552495	Homo sapiens	Human germline T-cell receptor beta chain Dopamine-beta-hydroxylase-like, TRY1, TRY2, TRY3, TCRBV27S1P, TCRBV22S1A2N1T, TCRBV9S1A1T, TCRBV7S1A1N2T, TCRBV5S1A1T, TCRBV13S3, TCRBV6S7P, TCRBV7S3A2T, TCRBV13S2A1T, TCRBV9S2A2PT, TCRBV7S2A1N4T, TCRBV13S9/13S2A1T, TCRBV6S5A1N1, TCRBV30S1P, TCRBV31S1, TCRBV13S5, TCRBV6S1A1N1, TCRBV32S1P, TCRBV5S5P, TCRBV1S1A1N1, TCRBV12S2A1T, TCRBV21S1, TCRBV8S4P, TCRBV12S3, TCRBV21S3A2N2T, TCRBV8S5P, TCRBV13S1 genes from bases 1 to 267156 (section 1 of 3).	534	99
1256	gi3821867	Homo sapiens	T cell receptor beta chain variable region (BV22S1A2N1) mRNA, partial cds.	532	90
1257	gi8096269	Nicotiana tabacum	KED	320	25
1257	gi7549210	Babesia bigemina	200 kDa antigen p200	333	25
1257	gi6624128	Homo sapiens	PAC clone RP4-687K1 from 14, complete sequence.	288	28
1258	gi10440026	Homo sapiens	cDNA: FLJ23363 fis, clone HEP15507.	1438	100
1258	gi12805037	Homo sapiens	clone MGC:5306 IMAGE:3460203, mRNA, complete cds.	1438	100
1258	AAB43628	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1073.	1101	99
1259	gi3435244	Homo sapiens	centriole associated protein CEP110 mRNA, complete cds.	4915	99
1259	gi485858	Mus musculus	IB3/5-polypeptide	1728	55
1259	gi9916	Plasmodium	liver stage antigen	556	24

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		falciparum			
1260	gi7019815	Homo sapiens	cDNA FLJ20003 fis, clone ADKA01794.	812	52
1260	gi3878051	Caenorhabditis elegans	H21P03.2	179	32
1260	gi683703	Saccharomyces cerevisiae	QRI2	158	25
1261	gi14043196	Homo sapiens	clone IMAGE:3162799, mRNA, partial cds.	1247	100
1261	gi406738	Homo sapiens	H.sapiens SHB mRNA.	801	46
1261	AAY07040	Homo sapiens	Breast cancer associated antigen precursor sequence.	801	46
1262	gi10242353	Homo sapiens	pellino 2 (PELI2) mRNA, complete cds.	2290	100
1262	gi14550457	Homo sapiens	pellino (Drosophila) homolog 2, clone MGC:15066 IMAGE:3942712, mRNA, complete cds.	2290	100
1262	gi10242357	Mus musculus	pellino 2	2113	92
1263	gi212657	Gallus gallus	smooth muscle caldesmon	150	36
1263	gi7549210	Babesia bigemina	200 kDa antigen p200	147	26
1263	gi6599138	Homo sapiens	mRNA; cDNA DKFZp434I036 (from clone DKFZp434I036); partial cds.	147	26
1264	gi632549	Petromyzon marinus	NF-180	247	27
1264	gi9837381	Mus musculus	retinitis pigmentosa GTPase regulator	215	33
1264	gi703450	Ureaplasma parvum serovar 3	multiple banded antigen	211	28
1265	gi7595307	Homo sapiens	opioid growth factor receptor mRNA, complete cds.	82	28
1265	AAY92809	Homo sapiens	Human opioid growth factor receptor spliced version 8.	82	28
1265	gi7406950	Mus musculus	N system amino acids transporter NAT-1	79	32
1266	gi12751100	Homo sapiens	PNAS-127 mRNA, complete cds.	380	100
1266	gi4646197	Arabidopsis thaliana	T23K8.5	68	28
1266	gi13925364	Influenza A virus (A/Hong Kong/514/97(H5 N1))	RNA polymerase	57	27
1267	gi11611473	Mus musculus	Deltex3	1858	98
1267	AAY70439	Homo sapiens	Human Notch signalling protein, Deltex (hZDX)-3.	370	45
1267	AAY70440	Homo sapiens	Human Notch signalling protein, Deltex (hZDX)-4.	370	45
1268	gi14042300	Homo sapiens	cDNA FLJ14640 fis, clone NT2RP2001460, weakly similar to TRICHOHYALIN.	2667	99
1268	AAB95168	Homo sapiens	Human protein sequence SEQ ID NO:17221.	2667	99
1268	gi13278316	Mus musculus	Similar to RIKEN cDNA	2182	70

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			2610507L03 gene		
1269	gi6016842	Mus musculus	nuclear protein ZAP	2869	92
1269	AAY72168	Homo sapiens	Human RNA metabolism protein (RMEP-8).	2798	100
1269	gi887378	Homo sapiens	(clone zap3) mRNA, 3' end of cds.	1728	97
1270	gi12803769	Homo sapiens	clone MGC:3411 IMAGE:3629947, mRNA, complete cds.	660	100
1270	gi4689250	Homo sapiens	sorting nexin 5 (SNX5) mRNA, complete cds.	92	95
1270	gi7023288	Homo sapiens	cDNA FLJ10931 fis, clone OVARC1000564.	92	95
1271	AAG00714	Homo sapiens	Human secreted protein, SEQ ID NO: 4795.	527	99
1271	gi1732239	Mycoplasma fermentans	monocytic differentiation/activation factor	87	20
1271	gi4324459	Schizosaccharomyces pombe	caffeine-induced death protein 2	78	22
1272	gi2414626	Schizosaccharomyces pombe	ubiquitin regulatory domain (UBX) protein	368	30
1272	gi914992	Saccharomyces cerevisiae	Ydr330wp	314	35
1272	gi7527718	Arabidopsis thaliana	T5E21.7	341	37
1273	gi4263747	Homo sapiens	PAC clone RP4-771P4 from 7q11.21-q11.23, complete sequence.	729	79
1273	gi2827180	Homo sapiens	general transcription factor 2-I (GTF2I) mRNA, complete cds.	729	79
1273	gi2827203	Homo sapiens	general transcription factor 2-I (GTF2I) mRNA, alternatively spliced product, complete cds.	729	79
1274	gi14042066	Homo sapiens	cDNA FLJ14503 fis, clone NT2RM1000252, weakly similar to H.sapiens E-MAP-115 mRNA.	1397	75
1274	AAB94010	Homo sapiens	Human protein sequence SEQ ID NO:14130.	1397	75
1274	gi7581985	Homo sapiens	mRNA for E-MAP-115/105 (MAP gene).	421	34
1275	gi10439810	Homo sapiens	cDNA: FLJ23201 fis, clone KAIA38872.	1349	99
1275	gi13097606	Homo sapiens	clone MGC:10765 IMAGE:3606341, mRNA, complete cds.	927	97
1275	gi4495095	Mus musculus	erythroid cell-specific and testis-specific protein 1(ERT-1)	69	38
1276	gi602438	Bos taurus	phosphoprotein	960	91
1276	gi10436906	Homo sapiens	cDNA: FLJ20940 fis, clone ADSE01608.	886	100
1276	gi7243755	Homo sapiens	neuronal phosphoprotein DARPP-32 mRNA, partial cds.	508	100
1277	gi3811376	Mus musculus	Bing1	350	37
1277	gi4050099	Mus musculus	BING1	350	37
1277	gi2952308	Homo sapiens	zinc finger transcription factor (ZNF-X) mRNA, complete cds.	314	48
1278	gi14042529	Homo sapiens	cDNA FLJ14768 fis, clone NT2RP3004125, moderately similar	2742	99

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			to Mus musculus zinc finger protein splice variant FIZ1-B (Fiz1) mRNA.		
1278	AAB95788	Homo sapiens	Human protein sequence SEQ ID NO:18745.	2742	99
1278	gi5579307	Mus musculus	zinc finger protein splice variant FIZ1-A	2323	86
1279	gi15162128	Agrobacterium tumefaciens	AGR_pAT_493p	90	34
1279	gi1016116	Cyanophora paradoxa	ribosomal protein L1	81	25
1280	gi10434157	Homo sapiens	cDNA FLJ12582 fis, clone NT2RM4001151.	811	99
1280	AAB94144	Homo sapiens	Human protein sequence SEQ ID NO:14418.	811	99
1280	gi12653823	Homo sapiens	clone MGC:3101 IMAGE:3350198, mRNA, complete cds.	763	100
1281	AAB53801	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:1341.	307	95
1281	AAB99116	Homo sapiens	Human protein SEQ ID 14.	71	32
1281	AAB63454	Homo sapiens	Human breast cancer associated antigen protein sequence SEQ ID NO:816.	63	52
1282	gi14042423	Homo sapiens	cDNA FLJ14714 fis, clone NT2RP3001107, weakly similar to PEREGRIN.	2618	99
1282	AAB93095	Homo sapiens	Human protein sequence SEQ ID NO:11945.	2618	99
1282	gi6979019	Homo sapiens	BRL mRNA, complete cds.	527	33
1283	gi10432842	Homo sapiens	cDNA FLJ11561 fis, clone HEMBA1003142.	806	100
1283	AAB94927	Homo sapiens	Human protein sequence SEQ ID NO:16392.	806	100
1283	AAR10871	Homo sapiens	Exon III encoded by genomic meg-CSF clone.	62	30
1284	gi1752638	Homo sapiens	mRNA for RT14, complete cds.	630	100
1284	gi2769585	Mus musculus	GCN5L1 protein	603	96
1284	gi10803351	Ciona intestinalis	nuclear lamin	82	27
1285	AAY77471	Homo sapiens	Human deubiquitinating protein Dub11, SEQ ID NO:34.	1273	93
1285	AAW30711	Homo sapiens	Human ubiquitin-specific thiol protease DUB D38378.	1229	94
1285	AAY77470	Homo sapiens	Human deubiquitinating protein Dub11, SEQ ID NO:32.	1234	94
1286	gi10439995	Homo sapiens	cDNA: FLJ23342 fis, clone HEP13470.	2647	99
1286	gi12224847	Homo sapiens	mRNA; cDNA DKFZp667A213 (from clone DKFZp667A213).	1743	100
1286	gi13561988	Argiope trifasciata	major ampullate spidroin 2-like protein	123	42
1287	gi15080040	Homo sapiens	Similar to RIKEN cDNA 2310008J22 gene, clone MGC:19531 IMAGE:4336762, mRNA, complete cds.	708	46
1287	AAB95390	Homo sapiens	Human protein sequence SEQ ID NO:17738.	703	53

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1287	gi6599183	Homo sapiens	mRNA; cDNA DKFZp434P1514 (from clone DKFZp434P1514); partial cds.	687	49
1288	AAB95390	Homo sapiens	Human protein sequence SEQ ID NO:17738.	703	53
1288	gi15080040	Homo sapiens	Similar to RIKEN cDNA 2310008J22 gene, clone MGC:19531 IMAGE:4336762, mRNA, complete cds.	637	50
1288	gi6599183	Homo sapiens	mRNA; cDNA DKFZp434P1514 (from clone DKFZp434P1514); partial cds.	637	50
1289	gi10304393	Bombyx mori	ornithine decarboxylase antizyme	75	51
1289	gi331156	Human T-cell lymphotropic virus type 1	env polyprotein (partial)	51	45
1289	gi8980338	Takifugu rubripes	FRANK2 protein	76	48
1290	gi12751397	Homo sapiens	EVG1 mRNA, complete cds.	1106	100
1290	AAB64938	Homo sapiens	Human secreted protein sequence encoded by gene 3 SEQ ID NO:116.	87	33
1290	AAB64949	Homo sapiens	Human secreted protein sequence encoded by gene 10 SEQ ID NO:127.	87	33
1291	gi1263289	Araneus diadematus	fibroin-4	146	26
1291	AAY96125	Homo sapiens	Collagen type III alpha-1.	142	27
1291	gi930045	Homo sapiens	Human COL3A1 mRNA for pro alpha-1 (III) collagen.	142	27
1292	gi915208	Sus scrofa	gastric mucin	287	25
1292	gi13592175	Leishmania major	ppg3	292	26
1292	gi557822	Saccharomyces cerevisiae	mal5, sta1, len: 1367, CAI: 0.3, AMYH_YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	269	24
1293	AAB60458	Homo sapiens	Human cell cycle and proliferation protein CCYPR-6, SEQ ID NO:6.	923	100
1293	gi8920230	Homo sapiens	partial mRNA for Spir-1 protein (Spir-1 gene).	385	42
1293	gi4107015	Ciona savignyi	PEM-5	286	39
1294	gi13548492	Caenorhabditis elegans	ZC410.7b	354	44
1294	gi4996286	Arabidopsis thaliana	lipoyltransferase	343	40
1294	gi2494127	Arabidopsis thaliana	Contains similarity to Mycobacterium LIPB gene (gb Q104041).	343	40
1295	gi181097	Homo sapiens	Human gamma-A-crystallin gene (gamma-G5), exon 3.	968	99
1295	gi387135	Mus musculus	gamma-A-crystallin	845	84
1295	gi203627	Rattus norvegicus	gamma-A-crystallin	837	83
1296	AAG02474	Homo sapiens	Human secreted protein, SEQ ID NO: 6555.	266	96
1296	gi10518509	Streptomyces	B-N-acetylhexosaminidase	69	33

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		plicatus			
1296	gi7544050	Streptomyces coelicolor A3(2)	beta-N-acetylhexosaminidase	69	30
1297	gi7021900	Homo sapiens	cDNA FLJ10065 fis, clone HEMBA1001455.	981	38
1297	AAB92496	Homo sapiens	Human protein sequence SEQ ID NO:10598.	981	38
1297	gi6581093	Mus musculus	transposase-like protein	924	36
1298	gi12060820	Homo sapiens	serologically defined breast cancer antigen NY-BR-15 mRNA, partial cds.	2775	100
1298	gi10436303	Homo sapiens	cDNA FLJ13976 fis, clone Y79AA1001594, weakly similar to HYALURONAN-MEDIATED MOTILITY RECEPTOR.	858	37
1298	AAB95624	Homo sapiens	Human protein sequence SEQ ID NO:18344.	858	37
1299	gi12053099	Homo sapiens	mRNA; cDNA DKFZp434A171 (from clone DKFZp434A171); complete cds.	907	66
1299	AAB28628	Homo sapiens	Human B11Ag1 antigen splice isoform B11C-15.	596	49
1299	AAB28629	Homo sapiens	Human B11Ag1 antigen splice isoform B11C-8.	596	49
1300	gi12804657	Homo sapiens	clone IMAGE:3354845, mRNA, partial cds.	914	100
1300	AAG01406	Homo sapiens	Human secreted protein, SEQ ID NO: 5487.	421	100
1300	AAG00840	Homo sapiens	Human secreted protein, SEQ ID NO: 4921.	92	28
1301	gi13111480	Homo sapiens	mRNA for NeshBP, complete cds.	2167	84
1301	gi5912261	Homo sapiens	mRNA; cDNA DKFZp586L2024 (from clone DKFZp586L2024); partial cds.	1995	99
1301	AAE01791	Homo sapiens	Human gene 22 encoded secreted protein HOHDF66, SEQ ID NO:112.	970	70
1302	AAV34129	Homo sapiens	Human potassium channel K+Hnov28.	192	34
1302	AAB95201	Homo sapiens	Human protein sequence SEQ ID NO:17295.	192	34
1302	AAZ11907_aal	Homo sapiens	Human potassium channel K+Hnov28 cDNA (5' splice variant 1).	192	34
1303	AAB28049	Homo sapiens	Human secreted protein SEQ ID NO: 97.	204	100
1303	gi881373	Saccharomyces cerevisiae	ORF; Method: conceptual translation supplied by author	71	34
1303	gi927780	Saccharomyces cerevisiae	Ydr511wp; CAI: 0.13	71	34
1304	AAB50651	Homo sapiens	Human UNC-5H1 protein translation in frame 3 sequence SEQ ID NO:10.	93	32
1304	gi2232272	Schizophyllum commune	B2-aldehyde-forming enzyme	85	30

167
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1304	gi9800242	rat cytomegalovirus Maastricht	pr5	90	30
1305	gi10439606	Homo sapiens	cDNA: FLJ23042 fis, clone LNG02323.	955	100
1305	gi1041087	Mus musculus	synapsin II	85	32
1305	gi2460023	Homo sapiens	membrane-associated kinase (Myt1) mRNA, complete cds.	96	27
1306	gi1147611	Saccharomyces cerevisiae	Lpf3p	136	28
1306	gi1592074	Methanococcus jannaschii	tRNA intron endonuclease (endA)	93	32
1306	gi3093358	Mus musculus	SPR2A protein	74	40
1307	gi11225489	Homo sapiens	8q22.1 region and MTG8 (CBFA2T1) gene, partial cds.	94	57
1307	gi2981446	Homo sapiens	MTG8-related protein (MTGR2) gene, exon 10 and partial cds.	91	53
1307	gi10437096	Homo sapiens	cDNA: FLJ21080 fis, clone CAS02449.	108	51
1308	gi10433656	Homo sapiens	cDNA FLJ12229 fis, clone MAMMA1001181, weakly similar to ABC1 PROTEIN HOMOLOG PRECURSOR.	813	95
1308	AAB93936	Homo sapiens	Human protein sequence SEQ ID NO:13939.	813	95
1308	gi11595643	Neurospora crassa	probable abc1 protein precursor	144	40
1309	AAW30650	Homo sapiens	Human secreted protein clone bg249 1 protein.	514	36
1309	gi3878093	Caenorhabditis elegans	Similarity with drosophila MSP-300 protein (PIR acc. no. S30431), contains similarity to Pfam domain: PF01465 (GRIP domain), Score=90.2, E-value=1.4e-23, N=1	145	23
1309	gi4582571	Gallus gallus	Hyperion protein, 419 kD isoform	150	27
1310	gi4884134	Homo sapiens	mRNA; cDNA DKFZp586H0519 (from clone DKFZp586H0519); partial cds.	66	32
1310	gi12723703	Lactococcus lactis subsp. lactis	HYPOTHETICAL PROTEIN	61	34
1310	AAV70216	Homo sapiens	Partial human Interleukin-1 epsilon protein.	58	35
1311	gi3132823	Bos taurus	ribosomal protein L30	794	100
1311	gi560493	Rattus norvegicus	ribosomal protein L24	794	100
1311	gi12805289	Mus musculus	ribosomal protein L24	794	100
1312	gi10437683	Homo sapiens	cDNA: FLJ21562 fis, clone COL06420.	2858	99
1312	gi7899288	Homo sapiens	B2 gene partial cDNA, clone B2E.	398	34
1312	gi2076894	Caenorhabditis elegans	short region of weak similarity to protein kinase C; contains similarity to Pfam domain PF00130 (DAG_PE-bind), Score=10.0, E-value=0.0034, N=1	156	36

168
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1313	gi12053069	Homo sapiens	mRNA; cDNA DKFZp434J0617 (from clone DKFZp434J0617); complete cds.	1071	100
1313	gi296594	Hordeum vulgare	pZE40	89	35
1313	AAY84592	Homo sapiens	Amino acid sequence of a human artemin polypeptide.	116	34
1314	gi6599232	Homo sapiens	mRNA; cDNA DKFZp434M0728 (from clone DKFZp434M0728); partial cds.	1788	99
1314	gi10439715	Homo sapiens	cDNA: FLJ23129 fis, clone LNG08404.	1339	100
1314	gi927637	Anthocidaris crassispina	dynein intermediate chain 2	607	34
1315	gi11611545	Sus scrofa	aldose 1-epimerase	1670	89
1315	gi13422779	Caulobacter crescentus	aldose 1-epimerase	758	45
1315	gi9294498	Arabidopsis thaliana	aldose 1-epimerase-like protein	745	45
1316	gi5917666	Zea mays	extensin-like protein	154	25
1316	gi1749842	Yarrowia lipolytica	cell wall protein	123	35
1316	gi8163634	Streptococcus pneumoniae	surface protein PspC	125	25
1317	AAG01757	Homo sapiens	Human secreted protein, SEQ ID NO: 5838.	612	99
1317	gi1747	Oryctolagus cuniculus	trichohyalin	238	26
1317	gi13559600	Caenorhabditis elegans	short region of weak similarity to Plasmodium yoelii rhoptry protein (PIR:C45521)	218	24
1318	gi10438135	Homo sapiens	cDNA: FLJ21924 fis, clone HEP04086.	504	57
1318	gi15145793	Sus scrofa	basic proline-rich protein	317	38
1318	gi6523547	Volvox carteri f. nagariensis	hydroxyproline-rich glycoprotein DZ-HRGP	385	51
1319	gi14348588	Homo sapiens	KRAB zinc finger protein (KR18) mRNA, partial cds.	1269	60
1319	gi7023703	Homo sapiens	cDNA FLJ11191 fis, clone PLACE1007598, weakly similar to ZINC FINGER PROTEIN 184.	1389	62
1319	AAB93576	Homo sapiens	Human protein sequence SEQ ID NO:12988.	1389	62
1320	gi6467391	Homo sapiens	TONDU (TONDU) mRNA, complete cds.	160	37
1320	gi12652601	Homo sapiens	TONDU, clone MGC:2032 IMAGE:3504527, mRNA, complete cds.	160	37
1320	gi13097189	Homo sapiens	TONDU, clone MGC:5266 IMAGE:2900293, mRNA, complete cds.	160	37
1321	gi10435586	Homo sapiens	cDNA FLJ13544 fis, clone PLACE1006815.	1399	100
1321	AAB95456	Homo sapiens	Human protein sequence SEQ ID NO:17927.	1399	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1321	gi4761481	Homo sapiens	isolate 27 immunoglobulin lambda light chain variable region (IGL) gene, partial cds.	82	36
1322	gi10435982	Homo sapiens	cDNA FLJ13841 fis, clone THYRO1000787.	3838	99
1322	AAB94774	Homo sapiens	Human protein sequence SEQ ID NO:15862.	3838	99
1322	gi14336702	Homo sapiens	16p13.3 sequence section 3 of 8.	283	26
1323	gi9885296	Homo sapiens	LENG1 protein (LENG1) mRNA, partial cds.	1056	100
1323	AAB54164	Homo sapiens	Human pancreatic cancer antigen protein sequence SEQ ID NO:616.	823	98
1323	AAB43890	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1335.	102	24
1324	gi57720	Rattus rattus	ribosomal protein S20 (AA 1-119)	331	100
1324	gi15030141	Mus musculus	Similar to ribosomal protein S20	331	100
1324	gi13960133	Homo sapiens	ribosomal protein S20, clone MGC:4151 IMAGE:3029762, mRNA, complete cds.	331	100
1325	gi1345408	Mus musculus	AT motif-binding factor	453	40
1325	gi3924672	Homo sapiens	PAC clone RP5-991G20, complete sequence.	443	40
1325	AAR23963	Homo sapiens	AFP-I (Ala 2460 Val).	443	40
1326	gi12002024	Homo sapiens	brain my038 protein mRNA, complete cds.	371	100
1326	gi10178219	Arabidopsis thaliana	DNA-binding protein-like	100	41
1326	gi433638	Saccharomyces cerevisiae	D441	93	36
1327	gi6566147	Drosophila melanogaster	large Forked protein	200	32
1327	gi549986	Pennisetum ciliare	possible apospory-associated protein	156	34
1327	AAG02682	Homo sapiens	Human secreted protein, SEQ ID NO: 6763.	156	33
1328	gi10435867	Homo sapiens	cDNA FLJ13755 fis, clone PLACE3000363.	6782	99
1328	AAB94745	Homo sapiens	Human protein sequence SEQ ID NO:15792.	6782	99
1328	gi13277572	Homo sapiens	clone IMAGE:3456134, mRNA, partial cds.	2247	100
1329	gi14336699	Homo sapiens	16p13.3 sequence section 2 of 8.	832	51
1329	gi2961133	Drosophila melanogaster	nuclear fallout	338	28
1329	gi1353761	Naegleria fowleri	myosin II heavy chain	193	26
1330	gi15081398	Homo sapiens	kruppel-like zinc finger protein (ZNF300) mRNA, complete cds.	1901	58
1330	gi1020145	Homo sapiens	Human DNA binding protein (HPF2) mRNA, complete cds.	1469	57
1330	gi5679576	Homo sapiens	mRNA for zinc finger 41 (ZNF41 gene).	1596	47
1331	gi12804013	Homo sapiens	clone MGC:4179 IMAGE:3638050, mRNA, complete cds.	1351	99
1331	gi13938317	Homo sapiens	Similar to zinc finger protein 202,	389	69

170
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			clone MGC:15660 IMAGE:3347511, mRNA, complete cds.		
1331	gi3006231	Homo sapiens	PAC clone RP4-604G5 from 7q22-q31.1, complete sequence.	406	43
1332	gi56952	Rattus norvegicus	put. preoptic regulatory factor-2	381	98
1332	gi12052818	Homo sapiens	mRNA; cDNA DKFZp564B1162 (from clone DKFZp564B1162); complete cds.	232	34
1332	gi6226761	Dictyostelium discoideum	class VII unconventional myosin	201	28
1333	gi9309467	Mus musculus	leucine-rich glioma-inactivated 1 protein precursor	1128	46
1333	gi4091819	Homo sapiens	leucine-rich glioma-inactivated protein precursor (LGI1) mRNA, complete cds.	1121	46
1333	AAAY36299	Homo sapiens	Human secreted protein encoded by gene 76.	1111	45
1334	gi10434406	Homo sapiens	cDNA FLJ12735 fis, clone NT2RP2000258, weakly similar to ACTIVATOR 1 140 KD SUBUNIT.	4071	100
1334	AAB94270	Homo sapiens	Human protein sequence SEQ ID NO:14689.	4071	100
1334	gi167835	Dictyostelium discoideum	myosin heavy chain	233	19
1335	gi11493544	Homo sapiens	PRO1460	412	100
1335	gi6164848	Homo sapiens	transferrin receptor (TFRC) gene, complete cds.	77	39
1335	gi12654697	Homo sapiens	transferrin receptor (p90, CD71), clone MGC:3151 IMAGE:3354176, mRNA, complete cds.	77	39
1336	gi15079491	Homo sapiens	Similar to RIKEN cDNA 1200014H14 gene, clone MGC:20255 IMAGE:4651484, mRNA, complete cds.	1805	100
1336	gi13879544	Mus musculus	Similar to RIKEN cDNA 1200014H14 gene	934	88
1336	gi14043175	Homo sapiens	Similar to RIKEN cDNA 1200014H14 gene, clone IMAGE:3139657, mRNA, partial cds.	920	100
1337	AAB67449	Homo sapiens	Amino acid sequence of a human chaperone polypeptide.	566	100
1337	gi5824204	Schizosaccharomyces pombe	dnaj protein	158	29
1337	gi9757767	Arabidopsis thaliana	tetratricopeptide repeat protein 2-like	165	41
1338	gi5107082	Arabidopsis thaliana	small zinc finger-like protein	63	38
1338	gi3879008	Caenorhabditis elegans	Similarity to C.elegans mut-2 mutator strain transposon Tc5 transposase	85	22
1339	gi14024367	Mesorhizobium loti	weak similarity to ubiquinone/menaquinone	193	31

171
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			biosynthesis methyl transferase		
1339	gi4981950	Thermotoga maritima	ubiquinone/menaquinone biosynthesis methyltransferase-related protein	117	34
1339	gi1947037	Pantoea agglomerans	BioC	118	29
1340	gi559237	Petroselinum crispum	tyrosine-rich hydroxyproline-rich glycoprotein	97	25
1340	gi5917734	Cavia porcellus	sperad-7	74	35
1340	gi6090929	Rattus norvegicus	carbonyl reductase isoform I	49	37
1341	gi1710292	Homo sapiens	Human clone 23909 mRNA, partial cds.	1942	99
1341	AAW80398	Homo sapiens	A secreted protein encoded by clone cw1543_3.	1887	98
1341	gi7547029	Homo sapiens	GAP-like protein (N61) mRNA, complete cds.	298	36
1342	AAB54229	Homo sapiens	Human pancreatic cancer antigen protein sequence SEQ ID NO:681.	701	100
1342	AAB50867	Homo sapiens	Human zalpha31.	701	100
1342	AAW27638	Homo sapiens	Secreted protein AP162.	363	71
1343	gi13543979	Homo sapiens	clone MGC:13040 IMAGE:3622924, mRNA, complete cds.	1212	100
1344	AAY73346	Homo sapiens	HTRM clone 619699 protein sequence.	1338	60
1344	AAB43912	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1357.	1337	60
1344	gi12804721	Homo sapiens	clone MGC:2663 IMAGE:3543910, mRNA, complete cds.	1176	55
1345	gi13516895	Homo sapiens	MLZE mRNA, complete cds.	2166	100
1345	gi13516897	Mus musculus	MLZE	692	62
1345	gi10433559	Homo sapiens	cDNA FLJ12150 fis, clone MAMMA1000422.	255	37
1346	gi12053309	Homo sapiens	mRNA; cDNA DKFZp434I099 (from clone DKFZp434I099); complete cds.	1162	99
1346	gi6503225	Leishmania major	7138.7	142	36
1346	gi295941	Ovis aries	trichohyalin	115	22
1347	gi12751092	Homo sapiens	PNAS-123 mRNA, complete cds.	306	98
1347	gi9988480	Ursus maritimus	luteinizing hormone receptor	71	31
1347	gi296749	Macaca fascicularis	acrosin-trypsin inhibitor	60	37
1348	gi12018147	Chlamydomonas reinhardtii	vegetative cell wall protein gp1	216	30
1348	gi15145795	Sus scrofa	basic proline-rich protein	194	34
1348	gi10645308	Leishmania major	L8453.1	241	29
1349	gi14041991	Homo sapiens	cDNA FLJ14457 fis, clone HEMBB1002217, weakly similar to ZINC FINGER PROTEIN 91.	2187	100
1349	AAB95039	Homo sapiens	Human protein sequence SEQ ID NO:16797.	2187	100
1349	gi9963804	Homo sapiens	zinc finger protein ZNF286	1019	47

172
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			(ZNF286) mRNA, complete cds.		
1350	gi2429118	Leishmania major	DDI1; L3169.2	101	29
1350	gi7243252	Bombyx mori	protease-like protein	78	27
1350	gi14488325	Oryza sativa	Putative retroelement	84	33
1351	gi1799566	Mus musculus	stac	806	48
1351	gi1799568	Homo sapiens	mRNA for stac, complete cds.	796	46
1351	AAW59642	Homo sapiens	Amino acid sequence of human Stac protein.	796	46
1352	gi4691541	Homo sapiens	adenylate kinase 5 (AK5) mRNA, complete cds.	384	100
1352	AAW70495	Homo sapiens	Human disease related nucleotide kinase-3 (DRNK-3) protein sequence.	384	100
1352	gi8918488	Rattus norvegicus	adenylate kinase isozyme 1	361	36
1353	AAG01667	Homo sapiens	Human secreted protein, SEQ ID NO: 5748.	62	36
1353	AAAY86393	Homo sapiens	Human gene 20-encoded protein fragment, SEQ ID NO:308.	61	42
1353	gi222945	Rana catesbeiana	ANP precursor	73	22
1354	gi12654383	Homo sapiens	cytochrome c oxidase subunit VIb, clone MGC:3061 IMAGE:3344701, mRNA, complete cds.	259	54
1354	gi12803321	Homo sapiens	cytochrome c oxidase subunit VIb, clone MGC:1153 IMAGE:3347455, mRNA, complete cds.	259	54
1354	gi30295	Homo sapiens	Human mRNA for cytochrome c oxidase subunit VIb (EC 1.9.3.1).	259	54
1355	gi10434759	Homo sapiens	cDNA FLJ12969 fis, clone NT2RP2005841, weakly similar to Homo sapiens mRNA for ALEX3.	534	29
1355	AAB95263	Homo sapiens	Human protein sequence SEQ ID NO:17448.	534	29
1355	gi15029763	Mus musculus	Similar to RIKEN cDNA 1200004E24 gene	316	29
1356	gi57121	Rattus norvegicus	ribosomal protein L37	510	96
1356	gi461232	Homo sapiens	Human mRNA for ribosomal protein L37, complete cds.	510	96
1356	gi292441	Homo sapiens	ribosomal protein L37 mRNA, complete cds.	510	96
1357	gi14329698	Homo sapiens	partial mRNA for Doublesex-mab-3 (DM) domain, (DMRTC1 gene).	326	100
1357	gi14329696	Homo sapiens	partial mRNA for Doublesex-mab-3 (DM) domain, (DMRTC2 gene).	323	47
1357	gi385240	Sus scrofa	small proline-rich protein	73	32
1358	gi4056562	Homo sapiens	TNNT1 gene, exons 1-11 (and joined CDS).	1205	100
1358	gi339783	Homo sapiens	Human slow skeletal muscle troponin T mRNA, clone M1.	1205	100
1358	gi546023	human, skeletal and cardiac muscle, mRNA, 899 nt]. [Homo	troponin T slow isoform {alternatively spliced}	1205	100

173
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		sapiens			
1359	gi13272546	Physarum polycephalum	major plasmodial myosin heavy chain	186	22
1359	gi553309	Homo sapiens	Human (clone SY2/10) golgin-165 mRNA, partial cds.	176	20
1359	AAB69070	Homo sapiens	Human male enhanced antigen-2 (MEA-2) protein sequence SEQ ID NO:2.	176	20
1360	gi14456079	Drosophila melanogaster	fucosyltransferase homologue	709	38
1360	gi13436134	Homo sapiens	clone MGC:11141 IMAGE:3837094, mRNA, complete cds.	555	92
1360	gi14456077	Drosophila melanogaster	alpha 1,3-fucosyltransferase	283	33
1361	gi14529886	Mus musculus	bM145O4.1 (novel protein)	1090	53
1361	gi10439066	Homo sapiens	cDNA: FLJ22612 fis, clone HSI04965.	1096	56
1361	AAE02779	Homo sapiens	Human PRO-C-MG.45 protein encoded by DNA-C-MG.45-1776 cDNA clone.	1079	47
1362	gi2323287	multiple sclerosis associated retrovirus	polyprotein	3109	82
1362	gi535518	Murine leukemia virus	gag-pol polyprotein	1292	39
1362	gi331995	AKV murine leukemia virus	gag-pol polyprotein (tag amber codon at 2250-2252 inserts Gln in Mo-MuLV)	1291	40
1363	gi10439942	Homo sapiens	cDNA: FLJ23305 fis, clone HEP11392.	3521	99
1363	gi14041889	Homo sapiens	cDNA FLJ14394 fis, clone HEMBA1003235, weakly similar to TROPOMYOSIN.	1062	100
1363	AAB92554	Homo sapiens	Human protein sequence SEQ ID NO:10741.	1062	100
1364	gi5738559	Homo sapiens	mRNA for zinc finger protein, clone cZNF41.8, partial.	324	64
1364	gi5679576	Homo sapiens	mRNA for zinc finger 41 (ZNF41 gene).	324	64
1364	gi5738553	Homo sapiens	mRNA for zinc finger protein, clone cZNF41.5, partial.	276	65
1365	gi7020645	Homo sapiens	cDNA FLJ20500 fis, clone KAT09159.	274	35
1365	gi12052860	Homo sapiens	mRNA; cDNA DKFZp564O2071 (from clone DKFZp564O2071); complete cds.	274	35
1365	AAB51661	Homo sapiens	Human secreted protein sequence encoded by gene 42 SEQ ID NO:101.	274	35
1366	gi6599235	Homo sapiens	mRNA; cDNA DKFZp434E2321 (from clone DKFZp434E2321); partial cds.	1410	94
1366	gi14029609	Homo sapiens	nuclear receptor transcription	84	35

174
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			cofactor (SHARP) mRNA, complete cds.		
1366	gi205682	Rattus norvegicus	heavy neurofilament polypeptide	98	22
1367	gi6409379	Homo sapiens	zinc finger protein ZNF229 (ZNF229) mRNA, partial cds.	2299	100
1367	gi10434850	Homo sapiens	cDNA FLJ13029 fis, clone NT2RP3001057, moderately similar to ZINC FINGER PROTEIN 91.	1941	48
1367	AAB95278	Homo sapiens	Human protein sequence SEQ ID NO:17486.	1941	48
1368	AAG04049	Homo sapiens	Human secreted protein, SEQ ID NO: 8130.	490	90
1368	gi14532516	Arabidopsis thaliana	AT5g05210/K2A11_8	98	23
1368	gi161292	Loligo pealei	neurofilament protein	101	21
1369	gi13182797	Homo sapiens	leucine-rich-repeat protein RNO2 mRNA, complete cds.	1353	97
1369	gi10198209	Homo sapiens	caspase recruitment domain protein 7 mRNA, complete cds.	911	40
1369	gi5911939	Homo sapiens	mRNA; cDNA DKFZp586O1822 (from clone DKFZp586O1822); partial cds.	908	43
1370	gi13183793	Homo sapiens	CECR2 protein (CECR2) mRNA, complete cds.	5992	95
1370	gi1502355	Saccharomyces cerevisiae	GCN5	248	33
1370	gi9931486	Mus musculus	cell proliferation related protein CAP	206	30
1371	gi10800858	Homo sapiens	mRNA for aminopeptidase B (RNPEP gene).	3479	99
1371	gi15082509	Homo sapiens	arginyl aminopeptidase (aminopeptidase B), clone MGC:20408 IMAGE:4636934, mRNA, complete cds.	3435	98
1371	gi10933784	Homo sapiens	partial mRNA for aminopeptidase B (APB gene).	3431	98
1372	gi12804031	Homo sapiens	clone MGC:10433 IMAGE:3941786, mRNA, complete cds.	2355	100
1372	AAB64858	Homo sapiens	Human secreted protein sequence encoded by gene 41 SEQ ID NO:144.	774	100
1372	gi7332094	Caenorhabditis elegans	contains similarity to Pfam family PF00076 (RNA recognition motif), score=82.2, E=1e-20, N=1	500	47
1373	gi6841246	Homo sapiens	HSPC298	528	90
1373	gi9758695	Arabidopsis thaliana	antifungal protein-like	57	36
1373	gi12006106	Homo sapiens	IRA2 mRNA, partial cds, alternatively spliced.	54	57
1374	gi6984160	Streptococcus cristatus	srpA	205	17
1374	gi532113	Caenorhabditis elegans	homeotic region most like HMPB_DROME: homeotic	196	20

175

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			proboscipedia protein		
1374	gi161292	Loligo pealei	neurofilament protein	181	22
1375	gi14249827	Homo sapiens	clone MGC:10992 IMAGE:3637387, mRNA, complete cds.	1714	100
1375	gi14602889	Homo sapiens	clone MGC:13119 IMAGE:4100726, mRNA, complete cds.	1714	100
1375	gi13279188	Homo sapiens	clone IMAGE:3626861, mRNA, partial cds.	1354	100
1376	gi10437767	Homo sapiens	cDNA: FLJ21628 fis, clone COL08076.	2196	100
1376	gi9502202	Homo sapiens	endothelial zinc finger protein induced by tumor necrosis factor alpha (EZFIT) mRNA, complete cds.	1734	65
1376	gi488555	Homo sapiens	Human zinc finger protein ZNF135 mRNA, complete cds.	1473	58
1377	gi13543446	Homo sapiens	clone IMAGE:3529534, mRNA, partial cds.	4267	99
1377	gi1915977	Homo sapiens	Human CSF-1 receptor (FMS) gene, complete cds, and (SMF) gene, partial cds.	1994	100
1377	gi557822	Saccharomyces cerevisiae	mal5, sta1, len: 1367, CAI: 0.3, AMYH_YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	160	21
1378	gi12053173	Homo sapiens	mRNA; cDNA DKFZp434A112 (from clone DKFZp434A112); complete cds.	3184	83
1378	gi10434445	Homo sapiens	cDNA FLJ12761 fis, clone NT2RP2001378, weakly similar to MUCIN 2 PRECURSOR.	3143	99
1378	AAB94287	Homo sapiens	Human protein sequence SEQ ID NO:14728.	3143	99
1379	gi5821143	Homo sapiens	mRNA for RNA binding protein, partial cds, clone: R03.	156	27
1379	gi6649242	Homo sapiens	splicing coactivator subunit SRm300 (SRM300) mRNA, complete cds.	156	27
1379	gi18734	Glycine max	DNA-directed RNA polymerase	153	32
1380	gi847724	Homo sapiens	Human methylthioadenosine phosphorylase (MTAP) mRNA, complete cds.	1457	95
1380	gi11602392	Homo sapiens	methylthioadenosine phosphorylase (MTAP) mRNA, complete cds.	1454	95
1380	gi13111476	Mus musculus	methylthioadenosine phosphorylase	1370	90
1381	gi12584839	Homo sapiens	HT036-ISO (HT036-ISO) mRNA, complete cds.	826	94
1381	gi12584841	Homo sapiens	HT036 (HT036) mRNA, complete cds.	826	94
1381	AAG03594	Homo sapiens	Human secreted protein, SEQ ID NO: 7675.	398	88
1382	gi12584839	Homo sapiens	HT036-ISO (HT036-ISO) mRNA, complete cds.	931	91
1382	gi12584841	Homo sapiens	HT036 (HT036) mRNA, complete cds.	754	89

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1382	gi12543184	Corynebacterium glutamicum	RXA01089	312	43
1383	gi12584839	Homo sapiens	HT036-ISO (HT036-ISO) mRNA, complete cds.	916	99
1383	gi12584841	Homo sapiens	HT036 (HT036) mRNA, complete cds.	739	99
1383	gi146120	Escherichia coli	glyoxylate-induced protein	324	39
1384	gi10440340	Homo sapiens	cDNA: FLJ23598 fis, clone LNG15381.	330	82
1384	gi10434023	Homo sapiens	cDNA FLJ12500 fis, clone NT2RM2001675.	292	53
1384	AAB94072	Homo sapiens	Human protein sequence SEQ ID NO:14260.	292	53
1385	AAB95751	Homo sapiens	Human protein sequence SEQ ID NO:18660.	155	36
1385	AAAY16782	Homo sapiens	Human secreted protein (clone cg426_8).	141	54
1385	gi12053181	Homo sapiens	mRNA; cDNA DKFZp434F142 (from clone DKFZp434F142); complete cds.	99	35
1386	AAB92536	Homo sapiens	Human protein sequence SEQ ID NO:10701.	2682	100
1386	AAW29657	Homo sapiens	Human secreted protein BP101_2 (alternative).	1615	100
1386	AAW29650	Homo sapiens	Human secreted protein BP101_2.	1579	100
1387	gi6103649	Homo sapiens	F-box protein FBX10 mRNA, partial cds.	2878	99
1387	gi6164739	Homo sapiens	F-box protein Fbx10 (FBX10) mRNA, partial cds.	963	100
1387	AAAY83080	Homo sapiens	F-box protein FBP-12.	963	100
1388	gi15082426	Homo sapiens	Similar to RIKEN cDNA 2810055F11 gene, clone MGC:20203 IMAGE:4684687, mRNA, complete cds.	1840	99
1388	gi13435795	Mus musculus	Similar to RIKEN cDNA 2810055F11 gene	1661	88
1388	gi9695312	Trypanosoma cruzi	B-cell mitogen precursor	483	32
1389	gi1017722	Homo sapiens	Human repressor transcriptional factor (ZNF85) mRNA, complete cds.	2077	73
1389	gi186774	Homo sapiens	Human Kruppel related zinc finger protein (HTF10) mRNA, complete cds.	2068	67
1389	gi6088100	Homo sapiens	mRNA for zinc finger protein (ZFD25), complete cds.	2029	68
1390	gi10434854	Homo sapiens	cDNA FLJ13031 fis, clone NT2RP3001113, weakly similar to INTRACELLULAR PROTEIN TRANSPORT PROTEIN USO1.	3194	100
1390	AAB95279	Homo sapiens	Human protein sequence SEQ ID NO:17488.	3194	100
1390	gi3044185	Plasmodium falciparum	mature parasite-infected erythrocyte surface antigen	171	18
1391	gi6561965	Arabidopsis	kinesin-like protein	636	48

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		thaliana			
1391	gi12862607	Mus musculus	kinesin superfamily protein 19A	664	94
1391	gi2239242	Schizosaccharomyces pombe	kinesin-like protein	644	49
1392	gi12053319	Homo sapiens	mRNA; cDNA DKFZp586I0418 (from clone DKFZp586I0418); complete cds.	1011	99
1392	gi10436265	Homo sapiens	cDNA FLJ13954 fis, clone Y79AA1001145.	401	98
1392	AAB94842	Homo sapiens	Human protein sequence SEQ ID NO:16016.	401	98
1393	gi13442786	Mus musculus	Drctnbl1a	773	70
1393	gi13442784	Homo sapiens	mRNA for DRCTNNB1A, complete cds.	752	66
1393	gi2822468	Poeciliopsis occidentalis occidentalis	MHC class II protein	65	33
1394	gi12653235	Homo sapiens	eukaryotic translation initiation factor 3, subunit 3 (gamma, 40kD), clone MGC:8431 IMAGE:2821133, mRNA, complete cds.	1773	95
1394	gi2351380	Homo sapiens	translation initiation factor eIF3 p40 subunit mRNA, complete cds.	1773	95
1394	gi3986482	Homo sapiens	translation initiation factor eIF3 p40 subunit gene, exon 8 and complete cds.	1769	95
1395	gi3169261	Mus musculus	T-box transcription factor	1584	98
1395	gi3128382	Mus musculus	MmTbx14	1398	98
1395	gi12082748	Mus musculus	T-box transcription factor TBX18	970	87
1396	gi10435100	Homo sapiens	cDNA FLJ13189 fis, clone NT2RP3004253.	702	99
1396	AAB95361	Homo sapiens	Human protein sequence SEQ ID NO:17667.	702	99
1396	gi6601502	Arabidopsis thaliana	arginine/serine-rich protein	104	30
1397	gi12655165	Homo sapiens	zinc finger protein 256, clone MGC:1413 IMAGE:3138611, mRNA, complete cds.	2656	100
1397	AAZ55698_aa1	Homo sapiens	cDNA encoding human zinc finger protein BMZF3.	2647	99
1397	gi4894364	Homo sapiens	zinc finger protein 3	2646	99
1398	gi458692	Homo sapiens	Human (H326) mRNA, complete cds.	2060	67
1398	gi200241	Mus musculus	protein PC326	1597	59
1398	AAB57007	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1585.	763	74
1399	gi6467391	Homo sapiens	TONDU (TONDU) mRNA, complete cds.	158	66
1399	gi12652601	Homo sapiens	TONDU, clone MGC:2032 IMAGE:3504527, mRNA, complete cds.	158	66
1399	gi13097189	Homo sapiens	TONDU, clone MGC:5266 IMAGE:2900293, mRNA, complete cds.	158	66
1400	gi14335444	Homo sapiens	axonemal dynein heavy chain 8	4347	70

178
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			(DNAH8) mRNA, complete cds.		
1400	gi14335446	Mus musculus	axonemal dynein heavy chain 8 long form	4347	70
1400	gi557716	Chlamydomonas reinhardtii	gamma heavy chain subunit of outer-arm dynein	2807	47
1401	gi14198333	Homo sapiens	ribosomal protein L12, clone MGC:9760 IMAGE:3855674, mRNA, complete cds.	754	92
1401	gi186800	Homo sapiens	Human ribosomal protein L12 mRNA, complete cds.	754	92
1401	AAG03810	Homo sapiens	Human secreted protein, SEQ ID NO: 7891.	754	92
1402	gi7020881	Homo sapiens	cDNA FLJ20643 fis, clone KAT02633.	107	25
1402	AAV35946	Homo sapiens	Extended human secreted protein sequence, SEQ ID NO. 195.	99	24
1402	gi6467974	Dendrobium grex Madame Thong-In	MADS box protein DOMADS2	83	27
1403	gi15079625	Homo sapiens	clone IMAGE:3957606, mRNA, partial cds.	519	36
1403	gi12746532	Homo sapiens	DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA, complete cds.	143	33
1403	AAV54057	Homo sapiens	Amino acid sequence of novo DNA cytosine methyltransferase DNMT3A.	143	33
1404	gi13938351	Homo sapiens	Similar to zinc finger protein 268, clone IMAGE:3352268, mRNA, partial cds.	3615	99
1404	gi186774	Homo sapiens	Human Kruppel related zinc finger protein (HTF10) mRNA, complete cds.	2590	55
1404	gi10440398	Homo sapiens	mRNA for FLJ00032 protein, partial cds.	2209	57
1405	gi12053013	Homo sapiens	mRNA; cDNA DKFZp434J0113 (from clone DKFZp434J0113); complete cds.	4610	100
1405	gi1747	Oryctolagus cuniculus	trichohyalin	149	19
1405	gi9916	Plasmodium falciparum	liver stage antigen	146	22
1406	AAW42082	Homo sapiens	The amino acid sequence of the AS152_1 protein.	177	100
1406	AAW41936	Homo sapiens	Secreted protein AS152_1.	177	100
1406	gi157622	Drosophila melanogaster	hairless protein	65	45
1407	gi10435212	Homo sapiens	cDNA FLJ13262 fis, clone OVARC1000912.	1020	100
1407	AAB94541	Homo sapiens	Human protein sequence SEQ ID NO:15286.	1020	100
1407	gi12044832	Chlamydomonas reinhardtii	DEAH-box RNA helicase	132	34
1408	gi10834636	Homo sapiens	double FYVE-containing protein 1 mRNA, complete cds.	1874	100

179
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1408	gi13469801	Homo sapiens	tandem FYVE fingers-1 protein mRNA, complete cds.	1874	100
1408	gi14042048	Homo sapiens	cDNA FLJ14493 fis, clone MAMMA1002972.	1760	100
1409	gi12330993	Homo sapiens	testis-specific protein TSP-NY mRNA, complete cds.	2586	100
1409	gi29865	Homo sapiens	H.sapiens CENP-E mRNA.	149	26
1409	gi153727	Streptococcus sp. 'group G'	M protein	138	23
1410	gi12653703	Homo sapiens	clone MGC:3048 IMAGE:3343305, mRNA, complete cds.	1182	100
1410	gi476061	Saccharomyces cerevisiae	YBR0834	169	40
1410	gi3237301	Saccharomyces cerevisiae	Vid24p	169	40
1411	gi5669894	Human herpesvirus 8	latent nuclear antigen	232	22
1411	gi295941	Ovis aries	trichohyalin	183	24
1411	gi212449	Gallus gallus	nonmuscle myosin heavy chain	182	24
1412	gi12053319	Homo sapiens	mRNA; cDNA DKFZp586I0418 (from clone DKFZp586I0418); complete cds.	1195	100
1412	gi10436265	Homo sapiens	cDNA FLJ13954 fis, clone Y79AA1001145.	406	100
1412	AAB94842	Homo sapiens.	Human protein sequence SEQ ID NO:16016.	406	100
1413	gi15080358	Homo sapiens	Similar to LOC88745, clone MGC:20327 IMAGE:4300313, mRNA, complete cds.	1291	98
1413	gi13623395	Homo sapiens	clone IMAGE:4053965, mRNA, partial cds.	1265	98
1413	gi6841156	Homo sapiens	HSPC253	912	99
1414	gi15079971	Homo sapiens	Similar to FAST kinase, clone MGC:19784 IMAGE:3831196, mRNA, complete cds.	2275	92
1414	gi1006659	Homo sapiens	H.sapiens mRNA for FAST kinase.	2275	92
1414	AAR72830	Homo sapiens	Human TIABP2.	2275	92
1415	gi12654589	Homo sapiens	clone MGC:2550 IMAGE:2966683, mRNA, complete cds.	1252	100
1415	AAB73513	Homo sapiens	Human transferase HTFS-20, SEQ ID NO:20.	1252	100
1415	gi4200446	Mus musculus	FYVE finger-containing phosphoinositide kinase	205	33
1416	gi179331	Homo sapiens	B94 protein mRNA, complete cds.	403	27
1416	gi1163174	Rattus norvegicus	similar to yeast Sec6p, Swiss-Prot Accession Number P32844; similar to mammalian B94, Swiss-Prot Accession Number Q03169; Method: conceptual translation supplied by author	246	24
1416	AAAY51115	Homo sapiens	Human HSEC6 protein.	229	24
1417	gi10440416	Homo sapiens	mRNA for FLJ00043 protein, partial cds.	6737	95
1417	gi13604143	Mus musculus	tangerin A	1969	50
1417	gi13604145	Mus musculus	tangerin B	1155	79

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1418	gi4589375	Homo sapiens	mRNA for Gab2, complete cds.	438	43
1418	gi4589377	Mus musculus	Gab2	428	52
1418	gi4159804	Mus musculus	PH domain containing adaptor molecule p97/Gab2	428	52
1419	gi4589375	Homo sapiens	mRNA for Gab2, complete cds.	438	43
1419	gi4159804	Mus musculus	PH domain containing adaptor molecule p97/Gab2	428	52
1419	gi4589377	Mus musculus	Gab2	428	52
1420	gi13528684	Homo sapiens	Similar to ribosomal protein S6 kinase, 52kD, polypeptide 1, clone MGC:11287 IMAGE:3945902, mRNA, complete cds.	2737	100
1420	AAB65612	Homo sapiens	Novel protein kinase, SEQ ID NO: 138.	2065	100
1420	gi6466791	Homo sapiens	ribosomal S6 protein kinase mRNA, complete cds.	533	55
1421	gi2330003	Gallus gallus	glutamine rich protein	151	26
1421	gi211896	Gallus gallus	h-caldesmon	145	24
1421	gi600255	Gallus gallus	caldesmon	145	24
1422	gi12654493	Homo sapiens	clone MGC:2718 IMAGE:2821816, mRNA, complete cds.	1819	100
1422	AAY10843	Homo sapiens	Amino acid sequence of a human secreted protein.	84	36
1422	gi330135	human herpesvirus 1	latency-related protein 2	83	43
1423	gi2351200	Mus musculus	Ki antigen	1284	100
1423	gi3219288	Mus musculus	PA28 gamma subunit	1284	100
1423	gi12655139	Homo sapiens	Similar to proteasome (prosome, macropain) 28 subunit, 3, clone MGC:1394 IMAGE:3139753, mRNA, complete cds.	1284	100
1424	gi10176999	Arabidopsis thaliana	pectin acetyltransferase	170	25
1424	gi1431629	Vigna radiata	pectinacetyltransferase precursor	174	24
1424	gi3047082	Arabidopsis thaliana	similar to Vigna radiata pectinacetyltransferase precursor (GB:X99348)	171	29
1425	gi1817526	Anthocidaris crassispina	intermediate chain 1	412	58
1425	gi7580490	Homo sapiens	NM23-H8 (NME8) mRNA, complete cds.	480	40
1425	gi8308035	Danio rerio	nucleoside diphosphate kinase NDPK-Z6	227	39
1426	gi13097804	Homo sapiens	clone MGC:4368 IMAGE:2822930, mRNA, complete cds.	845	92
1426	gi13278804	Homo sapiens	clone MGC:2743 IMAGE:2822930, mRNA, complete cds.	845	92
1426	gi6580722	Yersinia pseudotuberculosis	O-unit polymerase-like protein	86	26
1427	gi3880445	Caenorhabditis elegans	contains similarity to Pfam domain: PF02214 (K+ channel tetramerisation domain), Score=79.5, E-value=2.3e-20, N=1	194	51
1427	AAY34129	Homo sapiens	Human potassium channel	201	43

181
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			K+Hnov28.		
1427	AAB95201	Homo sapiens	Human protein sequence SEQ ID NO:17295.	201	43
1428	gi9368881	Homo sapiens	mRNA; cDNA DKFZp547D065 (from clone DKFZp547D065).	1598	99
1428	AAE00674	Homo sapiens	Human protein tyrosine kinase receptor (PTK) from clone HSSJQ45.	1302	100
1428	gi6453590	Homo sapiens	mRNA; cDNA DKFZp434F2322 (from clone DKFZp434F2322).	1040	60
1429	AAB92801	Homo sapiens	Human protein sequence SEQ ID NO:11309.	1306	88
1429	gi12803785	Homo sapiens	Similar to cell cycle progression 2 protein, clone MGC:3620 IMAGE:3632005, mRNA, complete cds.	2341	100
1429	AAB94081	Homo sapiens	Human protein sequence SEQ ID NO:14279.	2334	99
1430	AAE01031	Homo sapiens	Human death domain-containing receptor (DDCR) protein from HFIHQ20 clone.	1259	100
1430	gi14091952	Rattus norvegicus	KIDINS220	396	39
1430	gi11321435	Rattus norvegicus	ankyrin repeat-rich membrane-spanning protein	394	39
1431	gi11875764	Homo sapiens	neurogenin 2 gene, partial cds.	1301	100
1431	gi1504095	Mus musculus	DNA-binding protein	1153	83
1431	gi1666910	Mus musculus	neurogenin 2	1145	83
1432	gi5230714	Pseudonaja textilis	long neurotoxin precursor	78	34
1432	AAB49653	Homo sapiens	Human SEC5 protein sequence SEQ ID 10.	50	52
1432	gi1628501	Vesicular stomatitis virus	phosphoprotein	68	35
1433	gi1881738	Acanthamoeba castellanii	myosin-I binding protein Acan125	391	29
1433	gi14701866	Dictyostelium discoideum	carmil	328	29
1433	gi13603845	Mus musculus	ribonuclease/angiogenin inhibitor 2	112	28
1434	gi12584159	Homo sapiens	zinc finger protein 268 (ZNF268) mRNA, complete cds.	5261	100
1434	gi14579579	Homo sapiens	ZNF268B mRNA, complete cds, alternatively spliced.	4395	100
1434	AAB93278	Homo sapiens	Human protein sequence SEQ ID NO:12322.	4006	98
1435	gi13094679	Mus musculus	ribosome receptor isoform mRRp47	245	26
1435	gi13094677	Mus musculus	ribosome receptor isoform mRRp61	234	24
1435	gi161958	Trypanosoma cruzi	surface antigen	227	21
1436	gi987652	Escherichia coli	hypothetical protein in UNG 3' region	129	26
1436	gi10638	Trypanosoma cruzi	ribosomal protein P-JL5	76	43
1436	gi436146	Trypanosoma cruzi	Tcp2beta	76	43

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1437	gi12653161	Homo sapiens	ribosomal protein L35, clone MGC:8582 IMAGE:2960987, mRNA, complete cds.	311	75
1437	gi562074	Homo sapiens	Human ribosomal protein L35 mRNA, complete cds.	311	75
1437	AAB59091	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 799.	311	75
1438	gi2282038	Homo sapiens	Arp2/3 protein complex subunit p21-Arc (ARC21) mRNA, complete cds.	457	96
1438	AAW29658	Homo sapiens	Homo sapiens BH272_3 clone secreted protein.	457	96
1438	gi2209347	Homo sapiens	p21-Arc mRNA, complete cds.	457	96
1439	gi14042550	Homo sapiens	cDNA FLJ14779 fis, clone NT2RP4000398, moderately similar to ZINC FINGER PROTEIN 140.	1428	65
1439	AAB93164	Homo sapiens	Human protein sequence SEQ ID NO:12091.	1428	65
1439	gi456269	Mus musculus domesticus	zinc finger protein 30	1349	53
1440	gi1322220	Homo sapiens	Nmi mRNA, complete cds.	91	32
1440	AAAY06223	Homo sapiens	Apoptin-associating Nmi/Hou-like protein.	91	32
1440	AAW73472	Homo sapiens	Human cancer-related protein Nmi.	91	32
1441	gi9294050	Arabidopsis thaliana	protein kinase-like protein	113	30
1441	gi12804025	Homo sapiens	clone MGC:11242 IMAGE:3939938, mRNA, complete cds.	100	38
1441	gi1914851	Mus musculus	WW domain binding protein 5; WBP5	62	58
1442	gi48556	Stenotrophomonas maltophilia	chorionic gonadotropin receptor homologue	63	38
1442	gi442514	Stenotrophomonas maltophilia	CG-like protein	63	38
1442	gi9758931	Arabidopsis thaliana	receptor protein kinase-like protein	60	28
1443	AAAY70847	Homo sapiens	Human NIP2b protein.	739	43
1443	gi12803291	Homo sapiens	BCL2/adenovirus E1B 19kD-interacting protein 2, clone MGC:1529 IMAGE:3343559, mRNA, complete cds.	718	46
1443	gi558844	Homo sapiens	BCL2/adenovirus E1B 19kD-interacting protein 2 (BNIP2) mRNA, complete cds.	718	46
1444	gi2429362	Santalum album	proline rich protein	150	34
1444	gi6456717	chayote mosaic tymovirus	overlapping protein	145	28
1444	gi10438473	Homo sapiens	cDNA: FLJ22184 fis, clone HRC00983.	141	27
1445	gi6911211	Mus musculus	ATP-binding cassette protein	268	71
1445	AAB95548	Homo sapiens	Human protein sequence SEQ ID NO:18171.	97	40
1445	gi3806158	Caenorhabditis	similar to the ATP-binding transport	101	39

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		elegans	protein family (ABC transporter) (Pfam: ABC_tran.hmm, score: 162.68 and 139.46)		
1446	gi633990	Mus musculus	zinc finger protein	1970	86
1446	gi4481920	Mus musculus	Ozrf1 protein	1967	86
1446	gi1336158	Rattus norvegicus	pancreas only zinc finger protein	1960	85
1447	gi13186114	Homo sapiens	mRNA for rab interacting lysosomal protein (RILP gene).	1577	94
1447	gi13436353	Homo sapiens	clone IMAGE:3632533, mRNA, partial cds.	1312	93
1447	gi7549210	Babesia bigemina	200 kDa antigen p200	156	30
1448	gi13186114	Homo sapiens	mRNA for rab interacting lysosomal protein (RILP gene).	2047	99
1448	gi13436353	Homo sapiens	clone IMAGE:3632533, mRNA, partial cds.	1782	100
1448	gi7549210	Babesia bigemina	200 kDa antigen p200	152	27
1449	AAE01693	Homo sapiens	Human gene 22 encoded secreted protein HOFNM53, SEQ ID NO:105.	320	67
1449	AAAY36238	Homo sapiens	Human secreted protein encoded by gene 15.	115	53
1449	AAB24023	Homo sapiens	Human PRO1410 protein sequence SEQ ID NO:9.	115	53
1450	gi12002226	Homo sapiens	C3HC4-type zinc finger protein (LZK1) mRNA, complete cds.	2042	99
1450	gi10437296	Homo sapiens	cDNA: FLJ21230 fis, clone COL00741.	2042	99
1450	gi13752272	Homo sapiens	laryngeal carcinoma related protein 1 mRNA, complete cds.	1504	99
1451	gi13752754	Homo sapiens	zinc finger 1111 mRNA, complete cds.	1951	62
1451	gi10436789	Homo sapiens	cDNA FLJ14345 fis, clone THYRO1001189, weakly similar to ZINC FINGER PROTEIN 91.	1873	60
1451	AAB95862	Homo sapiens	Human protein sequence SEQ ID NO:18929.	1873	60
1452	gi5262536	Homo sapiens	mRNA; cDNA DKFZp586G1822 (from clone DKFZp586G1822); partial cds.	72	35
1452	AAB53973	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:1513.	67	31
1452	gi1334643	Xenopus laevis	APEG precursor protein	76	34
1453	AAB63251	Homo sapiens	Human breast cancer associated antigen protein sequence SEQ ID NO:613.	964	99
1453	AAB63250	Homo sapiens	Human breast cancer associated antigen protein sequence SEQ ID NO:612.	415	74
1453	AAB63249	Homo sapiens	Human breast cancer associated antigen protein sequence SEQ ID NO:611.	381	81
1454	gi7239109	Homo sapiens	HSPC059	2117	63

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1454	gi13752754	Homo sapiens	zinc finger 1111 mRNA, complete cds.	2012	64
1454	gi186774	Homo sapiens	Human Kruppel related zinc finger protein (HTF10) mRNA, complete cds.	2004	62
1455	gi14042046	Homo sapiens	cDNA FLJ14492 fis, clone MAMMA1002937, weakly similar to ZINC FINGER PROTEIN 135.	262	28
1455	AAB95103	Homo sapiens	Human protein sequence SEQ ID NO:17076.	262	28
1455	gi14042143	Homo sapiens	cDNA FLJ14549 fis, clone NT2RM2001670, weakly similar to ZINC FINGER PROTEIN 29.	253	32
1456	gi15159614	Agrobacterium tumefaciens	AGR_L_2275p	191	37
1456	gi15076209	Sinorhizobium meliloti	HYPOTHETICAL PROTEIN	187	34
1456	gi7271117	Cicer arietinum	25.7 kDa protein	135	36
1457	gi14250716	Homo sapiens	clone MGC:13310 IMAGE:4110431, mRNA, complete cds.	1267	43
1457	AAB21018	Homo sapiens	Human nucleic acid-binding protein, NuABP-22.	1095	49
1457	AAB21042	Homo sapiens	Human nucleic acid-binding protein, NuABP-46.	662	54
1458	gi1110452	Homo sapiens	Human mRNA for hCREM (cyclic AMP-responsive element modulator) type1 alpha protein, complete cds.	594	99
1458	gi30217	Homo sapiens	H.sapiens mRNA for cAMP responsive-element modulator (CREM) (partial).	589	98
1458	gi7717221	Homo sapiens	cAMP responsive element modulator (CREM), exon G and partial cds.	403	93
1459	AAB54325	Homo sapiens	Human pancreatic cancer antigen protein sequence SEQ ID NO:777.	751	98
1459	gi14993576	Dictyostelium discoideum	deliriumA	226	29
1459	gi14488149	Homo sapiens	Authorin submission 'ENTRYNAME'.	195	33
1460	gi10436256	Homo sapiens	cDNA FLJ13949 fis, clone Y79AA1001041.	2706	100
1460	AAB94838	Homo sapiens	Human protein sequence SEQ ID NO:16008.	2706	100
1460	AAB67573	Homo sapiens	Amino acid sequence of a human hydrolytic enzyme HYENZ5.	2701	99
1461	AAG03605	Homo sapiens	Human secreted protein, SEQ ID NO: 7686.	299	98
1461	gi15077806	Danio rerio	zinc finger buttonhead-related transcription factor 1	55	39
1461	AAAY27742	Homo sapiens	Human secreted protein encoded by gene No. 36.	56	34
1462	gi11595643	Neurospora crassa	probable abc1 protein precursor	870	47

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1462	gi3087737	Arabidopsis thaliana	ABC1 protein	880	46
1462	gi1514641	Schizosaccharomyces pombe	abc1Sp	800	45
1463	gi14336679	Homo sapiens	16p13.3 sequence section 1 of 8.	1470	99
1463	gi8572544	Homo sapiens	3-methyl-adenine DNA glycosylase mRNA, complete cds.	1462	98
1463	AAB56919	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1497.	1459	98
1464	AAB56919	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1497.	1651	99
1464	gi8572544	Homo sapiens	3-methyl-adenine DNA glycosylase mRNA, complete cds.	1532	99
1464	gi14336679	Homo sapiens	16p13.3 sequence section 1 of 8.	1500	100
1466	gi14787176	Mus musculus	CSMD1	3011	92
1466	gi14787181	Homo sapiens	CUB and sushi multiple domains protein 1 short form mRNA, complete cds, alternatively spliced.	3046	95
1466	gi14794726	Homo sapiens	CUB and sushi multiple domains 1 protein mRNA, complete cds.	3046	95
1467	AAG02178	Homo sapiens	Human secreted protein, SEQ ID NO: 6259.	282	46
1467	gi470322	Daucus carota	proline-rich protein	108	31
1467	gi10198182	Cladrastis kentukea	ENOD2	105	28
1468	AAG01457	Homo sapiens	Human secreted protein, SEQ ID NO: 5538.	372	100
1468	gi7020037	Homo sapiens	cDNA FLJ20135 fis, clone COL06818.	194	37
1468	gi15209365	Xenopus laevis	kinesin-like protein	167	34
1469	gi4033608	Anthocidaris crassispina	B2HC	1476	42
1469	gi10710	Tripneustes gratilla	Beta heavy chain of outer-arm axonemal dynein ATPase	1189	35
1469	gi217203	Anthocidaris crassispina	dynein beta-heavy chain	1182	34
1470	gi8250039	Anopheles gambiae	serine protease-like protein	243	38
1470	gi15012124	Mus musculus	Similar to distal intestinal serine protease	243	47
1470	gi5777330	Homo sapiens	esp-1 mRNA for eosinophil serine protease, complete cds.	241	46
1471	gi9957760	Homo sapiens	KLK15 (KLK15) gene, complete cds, alternatively spliced.	1327	99
1471	gi14484922	Homo sapiens	prostinogen mRNA, complete cds.	1323	99
1471	gi11244759	Homo sapiens	serine protease gene cluster, complete sequence.	1230	94
1472	gi10435369	Homo sapiens	cDNA FLJ13373 fis, clone PLACE1000749.	2500	99
1472	AAB94615	Homo sapiens	Human protein sequence SEQ ID NO:15462.	2500	99
1472	AAV51846	Homo sapiens	Human 18.1 homolog protein fragment.	1382	100
1473	gi14043862	Homo sapiens	clone MGC:14138 IMAGE:3948518, mRNA, complete	1000	100

186
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			cds.		
1473	gi12053377	Homo sapiens	mRNA; cDNA DKFZp586C1924 (from clone DKFZp586C1924); complete cds.	1000	100
1473	AAG03905	Homo sapiens	Human secreted protein, SEQ ID NO: 7986.	545	100
1474	AAB47134	Homo sapiens	CDIFF-15, Incyte ID No. 3478571CD1.	1886	51
1474	gi7022185	Homo sapiens	cDNA FLJ10260 fis, clone HEMBB1000973, moderately similar to Mus musculus schlafen3 mRNA.	943	39
1474	AAB92636	Homo sapiens	Human protein sequence SEQ ID NO:10951.	943	39
1475	AAB47134	Homo sapiens	CDIFF-15, Incyte ID No. 3478571CD1.	1280	63
1475	gi7022185	Homo sapiens	cDNA FLJ10260 fis, clone HEMBB1000973, moderately similar to Mus musculus schlafen3 mRNA.	600	38
1475	AAB92636	Homo sapiens	Human protein sequence SEQ ID NO:10951.	600	38
1476	gi6708478	Mus musculus	formin-like protein	509	65
1476	gi4101720	Mus musculus	lymphocyte specific formin related protein	466	71
1476	gi13447468	Aspergillus nidulans	FH1/FH2 protein homolog	141	25
1477	AAG02288	Homo sapiens	Human secreted protein, SEQ ID NO: 6369.	561	100
1477	gi1495271	Arabidopsis thaliana	RNA helicase	290	58
1477	gi9293935	Arabidopsis thaliana	pre-mRNA splicing factor ATP-dependent RNA helicase-like protein	290	58
1478	AAV45013	Homo sapiens	Protein encoded by fchd531 gene.	443	40
1478	AAT94467_aa1	Homo sapiens	Human Fchd531 gene differentially regulated in endothelial cells.	443	40
1478	gi9886891	Mus musculus	zinc finger protein 276 C2H2 type	469	35
1479	gi6318855	Homo sapiens	A-kinase anchoring protein 220 mRNA, complete cds.	323	23
1479	gi1216411	Rattus norvegicus	AKAP 220	229	23
1479	gi8885520	Streptococcus gordonii	streptococcal hemagglutinin	200	20
1480	gi14042307	Homo sapiens	cDNA FLJ14644 fis, clone NT2RP2001756, weakly similar to ZINC FINGER PROTEIN 84.	1130	100
1480	AAB95176	Homo sapiens	Human protein sequence SEQ ID NO:17237.	1130	100
1480	AAB58718	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 426.	452	98
1481	AAG00431	Homo sapiens	Human secreted protein, SEQ ID NO: 4512.	232	80
1481	gi11120624	Homo sapiens	cationic trypsinogen (TRYP1) gene,	39	50

187
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			TRYP1-L104P allele, partial cds.		
1481	gi31863	Homo sapiens	H.sapiens gene for myeloperoxidase.	58	37
1482	gi14317890	Gallus gallus	spindlin	280	65
1482	gi12052957	Homo sapiens	mRNA; cDNA DKFZp566G0346 (from clone DKFZp566G0346); complete cds.	280	65
1482	gi14317888	Gallus gallus	spindlin	276	65
1483	gi3970964	Homo sapiens	PAC clone RP1-215P15 from 5q12, complete sequence.	1335	95
1483	gi3688111	Homo sapiens	P1 clone GSP13996 from 5q12, complete sequence.	1335	95
1483	gi1314346	Homo sapiens	Human survival motor neuron (SMN) gene, exons 7 and 8, and complete cds.	1335	95
1484	AAG02443	Homo sapiens	Human secreted protein, SEQ ID NO: 6524.	304	100
1484	AAAY50923	Homo sapiens	Human fetal brain cDNA clone vc10.1 derived protein.	69	43
1484	gi1679686	Homo sapiens	Human ataxin-2 related protein mRNA, partial cds.	68	35
1485	gi5817194	Homo sapiens	mRNA; cDNA DKFZp434F011 (from clone DKFZp434F011); partial cds.	408	100
1485	AAAY20298	Homo sapiens	Human apolipoprotein E mutant protein fragment 11.	104	38
1485	gi2130539	Homo sapiens	gallbladder mucin MUC5B mRNA, partial cds.	109	33
1486	AAW19919	Homo sapiens	Human Ksr-1 (kinase suppressor of Ras).	294	81
1486	gi1171250	Mus musculus	protein kinase related to Raf protein kinases; Method: conceptual translation supplied by author	279	78
1486	gi1322382	Cyprinus carpio	ZP2	175	43
1487	gi10439630	Homo sapiens	cDNA: FLJ23059 fis, clone LNG03912.	2810	99
1487	AAG03039	Homo sapiens	Human secreted protein, SEQ ID NO: 7120.	306	100
1487	gi4322670	Homo sapiens	dentin phosphoryn mRNA, complete cds.	116	18
1488	gi9454416	Mus musculus	zinc finger protein Sp5	2122	97
1488	gi15077806	Danio rerio	zinc finger buttonhead-related transcription factor 1	744	60
1488	gi466466	Rattus norvegicus	zinc finger protein	565	55
1489	gi4323152	Mus musculus	Ets-protein Spi-C	880	67
1489	gi8745414	Aulonocara hansbaenschi	Spi-C transcription factor	476	49
1489	gi11245502	Raja eglanteria	SpiC	449	42
1490	AAG01095	Homo sapiens	Human secreted protein, SEQ ID NO: 5176.	264	70
1490	gi7644350	Homo sapiens	golgi matrix protein GM130 (GOLGA2) mRNA, complete cds.	195	69
1490	gi8099669	Homo sapiens	golgin-like protein (GLP) mRNA, complete cds.	178	66
1491	gi12053013	Homo sapiens	mRNA; cDNA DKFZp434J0113	125	41

188
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			(from clone DKFZp434J0113); complete cds.		
1492	gi296881	Clostridium thermocellum	S-layer protein	156	31
1492	gi347377	Homo sapiens	Human MLL-AF4 der(11) fusion protein mRNA, complete cds.	106	22
1492	gi3822238	Lymantria dispar nucleopolyhedro virus	mucin-like protein	114	25
1493	gi6683745	Homo sapiens	HOTTTL protein mRNA, complete cds.	252	26
1493	gi5419859	Homo sapiens	mRNA; cDNA DKFZp434B103 (from clone DKFZp434B103); partial cds.	256	27
1493	AAB58909	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 617.	256	27
1494	gi14043433	Homo sapiens	clone IMAGE:3952677, mRNA, partial cds.	1371	100
1494	gi7023601	Homo sapiens	cDNA FLJ11127 fis, clone PLACE1006225.	590	41
1494	AAB93521	Homo sapiens	Human protein sequence SEQ ID NO:12861.	590	41
1495	gi7022422	Homo sapiens	cDNA FLJ10408 fis, clone NT2RM4000585.	1039	76
1495	AAB92790	Homo sapiens	Human protein sequence SEQ ID NO:11286.	1039	76
1495	gi15156995	Agrobacterium tumefaciens	AGR_C_3445p	125	27
1496	gi6567123	Mus musculus	mDj6	992	83
1496	gi3402485	Homo sapiens	mRNA for MRJ, complete cds.	771	62
1496	gi6681594	Homo sapiens	HSJ2 mRNA for DnaJ homolog, complete cds.	771	62
1497	gi8777659	Homo sapiens	transcription factor HOXD12 (HOXD12) and transcription factor HOXD11 (HOXD11) genes, complete cds.	1262	96
1497	gi871428	Mus musculus	Hox-4.6	1237	93
1497	gi397509	Mus musculus	HOXD-11	1173	93
1498	gi9916	Plasmodium falciparum	liver stage antigen	148	20
1498	gi7549210	Babesia bigemina	200 kDa antigen p200	130	21
1498	gi295941	Ovis aries	trichohyalin	108	20
1499	gi12803855	Homo sapiens	clone IMAGE:3616574, mRNA, partial cds.	1095	100
1499	gi6539606	Homo sapiens	metastasis suppressor protein mRNA, complete cds.	857	81
1499	gi1203820	Cricetinae gen. sp.	insulin receptor tyrosine kinase 53 kDa substrate	130	24
1500	gi12324510	Arabidopsis thaliana	AIG1-like protein; 11637-17773	195	28
1500	gi12324505	Arabidopsis thaliana	AIG1-like protein; 41133-42535	194	27
1500	gi12324503	Arabidopsis	AIG1-like protein; 37301-39301	166	29

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		thaliana			
1501	gi7384990	Rattus norvegicus	homeodomain protein RX	263	48
1501	gi12746271	Rattus norvegicus	retinal homeobox transcription factor Rx/rax	263	48
1501	gi2240024	Mus musculus	retinal homeobox protein	274	37
1502	gi2983243	Aquifex aeolicus	chromosome assembly protein homolog	103	24
1502	gi2764639	Mycoplasma hominis	Vaa surface lipoprotein adhesin precursor	96	22
1502	gi5457791	Pyrococcus abyssi	chromosome segregation protein (smc1)	97	21
1503	gi5262582	Homo sapiens	mRNA; cDNA DKFZp434K063 (from clone DKFZp434K063); partial cds.	3782	99
1503	AAE01527	Homo sapiens	Human gene 15 encoded secreted protein fragment, SEQ ID NO:184.	3358	100
1503	gi10438230	Homo sapiens	cDNA: FLJ21993 fis, clone HEP06576.	1416	100
1504	gi10434483	Homo sapiens	cDNA FLJ12788 fis, clone NT2RP2001946.	1492	99
1504	AAB95183	Homo sapiens	Human protein sequence SEQ ID NO:17252.	1492	99
1504	AAG01491	Homo sapiens	Human secreted protein, SEQ ID NO: 5572.	752	99
1505	gi10439485	Homo sapiens	cDNA: FLJ22944 fis, clone KAT08974.	1985	99
1505	gi12053299	Homo sapiens	mRNA; cDNA DKFZp434P078 (from clone DKFZp434P078); complete cds.	1566	100
1505	gi5532389	Lytechinus variegatus	microtubule-associated protein	96	24
1506	gi10435385	Homo sapiens	cDNA FLJ13385 fis, clone PLACE1001088.	1040	100
1506	AAB95409	Homo sapiens	Human protein sequence SEQ ID NO:17793.	1040	100
1506	gi6942015	Oryza sativa	extensin-like proline-rich protein RiP-5	60	24
1507	gi6094684	Homo sapiens	PAC clone RP1-278D1 from X, complete sequence.	984	34
1507	gi7023516	Homo sapiens	cDNA FLJ11078 fis, clone PLACE1005102, weakly similar to RING CANAL PROTEIN.	963	35
1507	AAB93480	Homo sapiens	Human protein sequence SEQ ID NO:12768.	963	35
1508	gi1478188	Gallus gallus	CHoxE	597	75
1508	gi6594621	Danio rerio	homeobox protein	314	57
1508	gi7689275	Xenopus laevis	homeodomain protein dbx	310	50
1509	AAG02366	Homo sapiens	Human secreted protein, SEQ ID NO: 6447.	391	98
1509	gi9965254	Bos taurus	RPGR-interacting protein 1 isoform b	107	42
1509	gi9966409	Bos taurus	RPGR-interacting protein-1	107	42
1510	gi496639	Mus musculus	transcription factor Ap-2 beta	1044	77
1510	gi3309577	Gallus gallus	transcription factor AP-2 beta	1039	77

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1510	gi1495417	Homo sapiens	H.sapiens mRNA for AP-2 beta transcription factor.	1039	77
1511	gi10436705	Homo sapiens	cDNA FLJ14280 fis, clone PLACE1005584, weakly similar to TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP0.	1429	100
1511	AAB95825	Homo sapiens	Human protein sequence SEQ ID NO:18834.	1429	100
1511	gi6015446	Hylobates pileatus	dopamine receptor D4	139	33
1512	gi10439972	Homo sapiens	cDNA: FLJ23325 fis, clone HEP12505.	3268	99
1512	gi10438214	Homo sapiens	cDNA: FLJ21982 fis, clone HEP06188.	2838	99
1512	AAB75305	Homo sapiens	Human secreted protein sequence encoded by gene 13 SEQ ID NO:124.	2785	98
1513	gi11275984	Homo sapiens	ubiquitin-like fusion protein mRNA, complete cds.	3802	99
1513	gi214866	Xenopus laevis	ubiquitin-like fusion protein	1098	46
1513	gi12964705	Rattus norvegicus	RSD-7	981	79
1514	gi15213542	Homo sapiens	NSD1 (NSD1) mRNA, complete cds.	11823	97
1514	gi3329465	Mus musculus	NSD1 protein	9872	82
1514	gi10438794	Homo sapiens	cDNA: FLJ22413 fis, clone HRC08475.	5795	100
1515	gi10436342	Homo sapiens	cDNA FLJ14001 fis, clone Y79AA1002298.	1404	99
1515	AAB94881	Homo sapiens	Human protein sequence SEQ ID NO:16098.	1404	99
1515	gi7022196	Homo sapiens	cDNA FLJ10267 fis, clone HEMBB1001056, weakly similar to PROLIFERATING-CELL NUCLEOLAR ANTIGEN P120.	100	30
1516	gi5360123	Homo sapiens	NY-REN-57 antigen mRNA, partial cds.	76	35
1516	gi6808184	Homo sapiens	mRNA; cDNA DKFZp434C0118 (from clone DKFZp434C0118); partial cds.	76	35
1516	AAAY92341	Homo sapiens	Human cancer associated antigen precursor from clone NY-REN-57.	76	35
1517	gi10434661	Homo sapiens	cDNA FLJ12903 fis, clone NT2RP2004364.	840	99
1517	AAB94367	Homo sapiens	Human protein sequence SEQ ID NO:14901.	840	99
1517	gi4322670	Homo sapiens	dentin phosphoryn mRNA, complete cds.	123	21
1518	gi6807686	Homo sapiens	mRNA; cDNA DKFZp434D0513 (from clone DKFZp434D0513).	5133	100
1518	gi10439597	Homo sapiens	cDNA: FLJ23035 fis, clone LNG02033.	1475	100
1518	gi10434608	Homo sapiens	cDNA FLJ12871 fis, clone NT2RP2003751.	107	30

191
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1519	gi10439957	Homo sapiens	cDNA: FLJ23315 fis, clone HEP12021.	3908	100
1519	gi10434127	Homo sapiens	cDNA FLJ12565 fis, clone NT2RM4000848.	3217	99
1519	AAB94133	Homo sapiens	Human protein sequence SEQ ID NO:14393.	3217	99
1520	gi10440377	Homo sapiens	mRNA for FLJ00024 protein, partial cds.	695	71
1520	gi1255702	Cicer arietinum	metallothionein	58	66
1521	gi13592175	Leishmania major	ppg3	145	24
1521	gi5420389	Leishmania major	proteophosphoglycan	127	25
1521	AAE03643	Homo sapiens	Human extracellular matrix and cell adhesion molecule-7 (XMAD-7).	113	25
1522	gi13177727	Homo sapiens	clone MGC:4606 IMAGE:3356565, mRNA, complete cds.	1293	96
1522	AAAY73335	Homo sapiens	HTRM clone 1850120 protein sequence.	1288	95
1522	gi1817526	Anthocidaris crassispina	intermediate chain 1	142	33
1523	gi529662	Homo sapiens	H.sapiens SCA1 mRNA for ataxin.	440	53
1523	AAAY33494	Homo sapiens	Human SCA1 protein.	440	53
1523	AAR71111	Homo sapiens	Spinocerebellar ataxia type 1 (SCA 1) protein product.	436	50
1524	AAG04066	Homo sapiens	Human secreted protein, SEQ ID NO: 8147.	414	97
1524	gi2760649	Leucania separata nuclear polyhedrosis virus	VP80	66	38
1524	gi2367597	Drosophila melanogaster	aryl hydrocarbon receptor nuclear translocator-like protein	61	32
1525	gi12659148	Mus musculus	mage-e2	656	88
1525	gi12659140	Mus musculus	mage-e1	478	43
1525	gi12659142	Mus musculus	mage-g1	401	39
1526	gi4336205	Zea mays	cytochrome b5 reductase	280	31
1526	gi14536592	Physcomitrella patens	PP001069030R	277	31
1526	gi14536588	Physcomitrella patens	25_ppprot1_046_e01	270	36
1527	gi10439267	Homo sapiens	cDNA: FLJ22757 fis, clone KAIA0803.	4111	100
1527	gi10437166	Homo sapiens	cDNA: FLJ21129 fis, clone CAS06266.	919	51
1527	gi10438855	Homo sapiens	cDNA: FLJ22457 fis, clone HRC09925.	306	26
1528	gi10439267	Homo sapiens	cDNA: FLJ22757 fis, clone KAIA0803.	4060	99
1528	gi10437166	Homo sapiens	cDNA: FLJ21129 fis, clone CAS06266.	921	51
1528	gi10438855	Homo sapiens	cDNA: FLJ22457 fis, clone HRC09925.	306	26
1529	gi7673675	Drosophila melanogaster	cactin	1041	45

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1529	gi4204304	Arabidopsis thaliana	lclprt_seq No definition line found	622	31
1529	AAY07071	Homo sapiens	Renal cancer associated antigen precursor sequence.	192	97
1530	gi300477	human, endogenous retroviral element RTVL-Hp3, Genomic Mutant, 240 nt, segment 1 of 2]. [Homo sapiens	pol=reverse transcriptase homolog {retroviral element}	305	75
1530	gi2323287	multiple sclerosis associated retrovirus	polyprotein	244	44
1530	gi6723273	Baboon endogenous virus strain M7	gag-pol precursor polyprotein	236	46
1531	gi8346962	Mus musculus	1A13 protein	1990	97
1531	gi7023188	Homo sapiens	cDNA FLJ10876 fis, clone NT2RP4001838, weakly similar to Homo sapiens CoREST protein mRNA.	1205	59
1531	AAB93285	Homo sapiens	Human protein sequence SEQ ID NO:12336.	1205	59
1532	gi2252814	Mus musculus	FOG	2128	70
1532	gi7595837	Xenopus laevis	Friend of GATA	1754	49
1532	gi5733391	Homo sapiens	zinc finger protein FOG-2 mRNA, complete cds.	514	37
1533	gi4079709	Rattus norvegicus	reggie1-1	2106	99
1533	gi4079711	Rattus norvegicus	reggie1-2	2063	96
1533	gi13277550	Homo sapiens	Similar to flotillin 2, clone MGC:5052, mRNA, complete cds.	1862	99
1534	gi14249875	Homo sapiens	clone IMAGE:3161752, mRNA, partial cds.	3062	98
1534	AAB93093	Homo sapiens	Human protein sequence SEQ ID NO:11941.	452	34
1534	AAB94716	Homo sapiens	Human protein sequence SEQ ID NO:15724.	299	41
1535	gi14249875	Homo sapiens	clone IMAGE:3161752, mRNA, partial cds.	3091	100
1535	AAB93093	Homo sapiens	Human protein sequence SEQ ID NO:11941.	480	35
1535	AAB94716	Homo sapiens	Human protein sequence SEQ ID NO:15724.	299	41
1536	gi7655	Drosophila melanogaster	betaH spectrin	127	28
1536	gi161334	Lytechinus variegatus	alpha-spectrin	120	28
1536	gi1495198	Rattus norvegicus	alphaII spectrin	122	29
1537	gi5669894	Human	latent nuclear antigen	127	24

193
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		herpesvirus 8			
1537	gi9438033	Homo sapiens	ser/arg-rich pre-mRNA splicing factor SR-A1 (SR-A1) gene, complete cds.	136	25
1537	gi10440402	Homo sapiens	mRNA for FLJ00034 protein, partial cds.	136	25
1538	gi12654495	Homo sapiens	clone IMAGE:2822295, mRNA, partial cds.	829	100
1538	gi552255	Lytechinus pictus	troponin C	146	51
1538	gi11191974	Bacillus circulans	BtrG	88	34
1539	gi10440109	Homo sapiens	cDNA: FLJ23425 fis, clone HEP22862.	3086	99
1539	gi5911984	Homo sapiens	mRNA; cDNA DKFZp434E0335 (from clone DKFZp434E0335); partial cds.	1764	100
1539	AAAY41528	Homo sapiens	Fragment of human secreted protein encoded by gene 77.	1172	99
1540	gi10435599	Homo sapiens	cDNA FLJ13556 fis, clone PLACE1007705, moderately similar to Mus musculus mRNA for Ndr1 related protein Ndr3.	1488	100
1540	AAB95462	Homo sapiens	Human protein sequence SEQ ID NO:17944.	1488	100
1540	gi12083721	Homo sapiens	mRNA for NDRG3, complete cds.	1246	99
1541	gi14089673	Mycoplasma pulmonis	50S RIBOSOMAL PROTEIN L35	229	74
1541	gi150150	Mycoplasma fermentans	ribosomal protein L35	213	69
1541	gi14247451	Staphylococcus aureus subsp. aureus Mu50	50S ribosomal protein L35	175	58
1542	gi13543686	Homo sapiens	Similar to RIKEN cDNA 4931428F02 gene, clone MGC:14797 IMAGE:4064169, mRNA, complete cds.	750	97
1542	AAAY07081	Homo sapiens	Renal cancer associated antigen precursor sequence.	267	38
1542	gi7020359	Homo sapiens	cDNA FLJ20333 fis, clone HEP11252.	206	31
1543	gi4007988	Mus musculus	D6MM5e protein	1651	74
1543	gi4008004	Mus musculus	D6MM5E protein	1651	74
1543	gi6010389	Enterobacteria phage Mu	gp15	64	26
1544	AAE03933	Homo sapiens	Human gene 36 encoded secreted protein HAUCC84, SEQ ID NO:96.	62	32
1544	AAE03976	Homo sapiens	Human gene 36 encoded secreted protein fragment, SEQ ID NO:168.	62	32
1544	gi1652299	Synechocystis sp. PCC 6803	DNA ligase	82	32
1545	gi13938461	Homo sapiens	Similar to RIKEN cDNA 4933428D01 gene, clone MGC:16186 IMAGE:3637472, mRNA, complete cds.	938	100

194
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1545	gi3688589	Triticum aestivum	MADS box transcription factor	94	30
1545	gi9367309	Hordeum vulgare	MADS-box protein 5	93	30
1546	gi10432716	Homo sapiens	cDNA FLJ11457 fis, clone HEMBA1001522.	816	58
1546	AAB94901	Homo sapiens	Human protein sequence SEQ ID NO:16290.	816	58
1546	gi13122230	Leishmania major	CDC27/NUC2-related protein	110	28
1547	gi14150461	Mus musculus	Osa1 nuclear protein	827	36
1547	gi5689365	Homo sapiens	gene for B120, exon 17 and complete cds.	817	36
1547	gi14133259	Homo sapiens	mRNA for SWI related protein, complete cds.	821	36
1548	gi7022971	Homo sapiens	cDNA FLJ10748 fis, clone NT2RP3001819, weakly similar to RING CANAL PROTEIN.	578	43
1548	AAB93127	Homo sapiens	Human protein sequence SEQ ID NO:12013.	578	43
1548	AAB92953	Homo sapiens	Human protein sequence SEQ ID NO:11635.	234	27
1549	gi10438282	Homo sapiens	cDNA: FLJ22031 fis, clone HEP08734.	1434	100
1549	gi9837427	Lytechinus variegatus	embryonic blastocoelar extracellular matrix protein precursor	278	29
1549	gi1002808	Lytechinus variegatus	extracellular matrix protein	152	25
1550	gi7288419	Treponema pallidum	tpr protein K	68	48
1550	gi11095629	Treponema pallidum	TprK	64	45
1550	gi7288427	Treponema pallidum	tpr protein K	63	45
1551	gi15026770	Clostridium acetobutylicum	Transcriptional regulator, MarR/EmrR family	57	42
1551	gi213301	Heterodontus francisci	H-chain V-region precursor	60	48
1551	gi63993	Heterodontus francisci	Ig heavy chain (Vh-Dh-Jh)	60	48
1552	gi12655191	Homo sapiens	glutathione transferase zeta 1 (maleylacetoacetate isomerase), clone MGC:2029 IMAGE:3139094, mRNA, complete cds.	529	89
1552	gi3510757	Homo sapiens	glutathione transferase zeta 1 (GSTZ1) gene, exon 9 and complete cds.	529	89
1552	gi2832731	Homo sapiens	mRNA for maleylacetoacetate isomerase.	529	89
1553	gi10435720	Homo sapiens	cDNA FLJ13646 fis, clone PLACE1011325.	858	100
1553	AAB94701	Homo sapiens	Human protein sequence SEQ ID NO:15686.	858	100
1553	gi9438179	Homo sapiens	inner centromere protein INCENP mRNA, complete cds.	186	26

195
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1554	gi9971158	Homo sapiens	GTPBP2 mRNA for GTP-binding like protein 2, complete cds.	1286	99
1554	gi9971156	Mus musculus	GTP-binding like protein 2	1257	98
1554	gi7020514	Homo sapiens	cDNA FLJ20423 fis, clone KAT02589.	1168	99
1555	gi15155438	Agrobacterium tumefaciens	AGR_C_894p	192	39
1555	gi3080529	Schizosaccharomyces pombe	alkB homolog-possible damage repair protein	80	27
1555	gi8163877	Entamoeba histolytica	enhancer binding protein-1; EBP1	101	27
1556	gi479171	Homo sapiens	H.sapiens nek2 mRNA for protein kinase.	1921	87
1556	gi507875	Homo sapiens	Human NIMA-like protein kinase 1 (NLK1) mRNA, complete cds.	1921	87
1556	AAAY92330	Homo sapiens	Human NIK1 protein.	1921	87
1557	gi479171	Homo sapiens	H.sapiens nek2 mRNA for protein kinase.	1948	89
1557	gi507875	Homo sapiens	Human NIMA-like protein kinase 1 (NLK1) mRNA, complete cds.	1948	89
1557	AAAY92330	Homo sapiens	Human NIK1 protein.	1948	89
1558	gi10438880	Homo sapiens	cDNA: FLJ22476 fis, clone HRC10682.	2605	100
1558	gi13469818	Mus musculus	bicaudal-C	2179	89
1558	gi7800180	Xenopus laevis	bicaudal-C	1720	73
1559	gi10436065	Homo sapiens	cDNA FLJ13893 fis, clone THYRO1001661.	1979	99
1559	AAB95608	Homo sapiens	Human protein sequence SEQ ID NO:18306.	1979	99
1559	gi10434804	Homo sapiens	cDNA FLJ12998 fis, clone NT2RP3000267.	1847	98
1560	AAB33840	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 184.	65	28
1560	gi13160459	Myzus persicae	cytochrome p450	64	33
1560	gi28435	Homo sapiens	Human mRNA for apoferritin H chain type.	79	24
1561	gi14336760	Homo sapiens	16p13.3 sequence section 7 of 8.	2672	75
1561	gi5869804	Drosophila melanogaster	cramped protein	274	29
1561	gi557822	Saccharomyces cerevisiae	mal5, sta1, len: 1367, CAI: 0.3, AMYH_YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	184	20
1562	gi10436504	Homo sapiens	cDNA FLJ14124 fis, clone MAMMA1002498.	896	100
1562	AAB95707	Homo sapiens	Human protein sequence SEQ ID NO:18550.	896	100
1562	gi15026974	Homo sapiens	mRNA for obscurin (OBSCN gene).	449	96
1563	gi5882290	Homo sapiens	Ras guanine nucleotide exchange factor 2 mRNA, partial cds.	6402	100
1563	gi2522208	Homo sapiens	Ras-GRF2 mRNA, partial cds.	6397	99
1563	gi57665	Rattus rattus	P140 RAS-GRF	4114	65
1564	gi5882290	Homo sapiens	Ras guanine nucleotide exchange factor 2 mRNA, partial cds.	6412	100
1564	gi2522208	Homo sapiens	Ras-GRF2 mRNA, partial cds.	6407	99
1564	gi57665	Rattus rattus	P140 RAS-GRF	4124	65

196
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1565	gi14043319	Homo sapiens	protein phosphatase 1, regulatory (inhibitor) subunit 2, clone MGC:1327 IMAGE:3346573, mRNA, complete cds.	503	79
1565	gi4704218	Homo sapiens	mRNA for inhibitor 2 of protein phosphatase 1.	503	79
1565	gi474388	Homo sapiens	H.sapiens mRNA for inhibitor 2 gene.	503	79
1566	gi3378116	Mus musculus	CC chemokine ABCD-1	75	25
1566	gi4585823	Mus musculus	CC chemokine DC/B-CK	75	25
1566	gi5757694	Mus musculus	macrophage-derived chemokine MDC/ABCD-1	75	25
1567	gi6808297	Homo sapiens	mRNA; cDNA DKFZp434B0610 (from clone DKFZp434B0610); partial cds.	997	100
1567	gi13676773	Chlamydomonas reinhardtii	PF6 protein	130	27
1567	gi14250426	Homo sapiens	clone IMAGE:3866238, mRNA, partial cds.	111	26
1568	gi10436412	Homo sapiens	cDNA FLJ14050 fis, clone HEMBA1006796.	1641	100
1568	AAB95872	Homo sapiens	Human protein sequence SEQ ID NO:18952.	1641	100
1568	gi2388676	Mytilus edulis	precollagen P	146	30
1569	gi13603867	Homo sapiens	ferritin heavy polypeptide-like 17 (FTHL17) mRNA, complete cds.	575	72
1569	gi211774	Gallus gallus	ferritin H subunit	527	63
1569	gi2369861	Gallus gallus	ferritin H chain	527	63
1570	gi14787176	Mus musculus	CSMD1	1576	68
1570	gi14787181	Homo sapiens	CUB and sushi multiple domains protein 1 short form mRNA, complete cds, alternatively spliced.	1571	68
1570	gi14794726	Homo sapiens	CUB and sushi multiple domains 1 protein mRNA, complete cds.	1571	68
1571	gi7019911	Homo sapiens	cDNA FLJ20059 fis, clone COL01349.	533	38
1571	gi10441750	Homo sapiens	kelch-like protein C3IP1 mRNA, complete cds.	447	35
1571	gi14042496	Homo sapiens	cDNA FLJ14750 fis, clone NT2RP3002948, weakly similar to RING CANAL PROTEIN.	447	35
1572	gi15145793	Sus scrofa	basic proline-rich protein	405	32
1572	gi5305335	Mycobacterium tuberculosis	proline-rich mucin homolog	386	30
1572	gi5917666	Zea mays	extensin-like protein	386	30
1573	gi10439536	Homo sapiens	cDNA: FLJ22986 fis, clone KAT11742.	456	96
1573	gi31303	Homo sapiens	H.sapiens fau mRNA.	456	96
1573	gi31305	Homo sapiens	H.sapiens fau 1 gene.	456	96
1574	gi10438123	Homo sapiens	cDNA: FLJ21916 fis, clone HEP03994.	1187	97
1574	gi6841176	Homo sapiens	HSPC263	550	46
1574	gi13960159	Homo sapiens	clone MGC:4584 IMAGE:3051266, mRNA, complete cds.	550	46
1575	gi2982673	Homo sapiens	mRNA for p27, complete cds.	1033	100

197
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1575	gi6469047	Mus musculus	C184S protein	925	89
1575	AAAY36056	Homo sapiens	Extended human secreted protein sequence, SEQ ID NO. 441.	707	91
1576	gi11611473	Mus musculus	Deltex3	517	51
1576	AAAY70440	Homo sapiens	Human Notch signalling protein, Deltex (hZDX)-4.	374	48
1576	AAAY70439	Homo sapiens	Human Notch signalling protein, Deltex (hZDX)-3.	374	48
1577	gi14042190	Homo sapiens	cDNA FLJ14576 fis, clone NT2RM4001092, weakly similar to ZINC FINGER PROTEIN GLO3.	2615	96
1577	AAB95129	Homo sapiens	Human protein sequence SEQ ID NO:17133.	2615	96
1577	gi9651765	Mus musculus	zinc finger protein 289	2450	90
1578	gi9622219	Rattus norvegicus	beta-catenin binding protein	2550	95
1578	gi7493836	Rattus norvegicus	multidomain presynaptic cytomatrix protein Piccolo	171	27
1578	gi1666689	Mus musculus	alpha-NAC, muscle-specific form gp220	174	25
1579	gi3868802	Mus musculus	c29	1708	74
1579	gi12642308	Ovis aries	type I keratin intermediate filament IRSa1	1668	72
1579	AAB56803	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1381.	1215	62
1580	gi6164743	Homo sapiens	F-box protein Fbx20 (FBX20) mRNA, partial cds.	1002	100
1580	AAAY83093	Homo sapiens	F-box protein FBP-25.	1002	100
1580	AAAY83072	Homo sapiens	F-box motif of FBP-25.	261	100
1581	gi2622381	Methanothermobacter thermautotrophicus	conserved protein	115	37
1581	gi5458479	Pyrococcus abyssi	HYDROXYACYLGLUTATHIONE HYDROLASE related (EC 3.1.2.6) (GLYOXALASE II)	84	26
1581	AAG04045	Homo sapiens	Human secreted protein, SEQ ID NO: 8126.	71	33
1582	gi12655474	Homo sapiens	mRNA for keratin associated protein 9.8 (KRTAP9.8 gene).	960	94
1582	gi12655470	Homo sapiens	mRNA for keratin associated protein 9.4 (KRTAP9.4 gene).	931	93
1582	gi12655466	Homo sapiens	mRNA for keratin associated protein 9.2 (KRTAP9.2 gene).	895	85
1583	gi515644	Homo sapiens	Human nucleotide-binding protein mRNA, complete cds.	1112	98
1583	gi6018193	Mus musculus	nucleotide-binding protein long form	1022	88
1583	AAAY73353	Homo sapiens	HTRM clone 1870914 protein sequence.	880	94
1584	gi10177622	Arabidopsis thaliana	gene_id:K6M13.11~	190	36
1584	gi499199	Schizosaccharomyces pombe	uvi22	157	33
1584	gi3184086	Schizosaccharomyces pombe	uv-induced protein uvi22	157	33

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		yces pombe			
1585	gi2072425	Homo sapiens	Human non-lens beta gamma-crystallin like protein (AIM1) mRNA, partial cds.	1066	40
1585	gi6807694	Homo sapiens	mRNA; cDNA DKFZp434L1713 (from clone DKFZp434L1713); partial cds.	1044	100
1585	gi162917	Bos taurus	gamma-B-crystallin	319	36
1586	AAG03361	Homo sapiens	Human secreted protein, SEQ ID NO: 7442.	204	97
1586	gi5706374	unidentified baculovirus	HHNBV-XIA	77	43
1586	gi13195754	Homo sapiens	protein tyrosine phosphatase mRNA, partial cds.	70	24
1587	gi3213225	Homo sapiens	T-box-containing transcriptional activator (TBX10) gene, exon 8, and partial cds.	953	97
1587	gi14289455	Homo sapiens	T-box 1 transcription factor C (TBX1C) mRNA, complete cds.	799	62
1587	gi12620817	Mus musculus	T-box 1	786	67
1588	AAAY17227	Homo sapiens	Human secreted protein (clone ya1-1).	100	25
1588	AAAY94938	Homo sapiens	Human secreted protein clone ye78_1 protein sequence SEQ ID NO:82.	93	30
1588	gi61709	Rous sarcoma virus	gag polyprotein; protein p19 (aa 1n84) (652 is 1st base in codon)	61	34
1589	gi10439883	Homo sapiens	cDNA: FLJ23259 fis, clone COL05779.	2480	99
1589	gi164639	Sus scrofa	ribonuclease inhibitor	269	30
1589	gi14577933	Mus musculus	ribonuclease/angiogenesis inhibitor	277	31
1590	gi13097207	Homo sapiens	ribosomal protein, large, P1, clone MGC:5215 IMAGE:2900846, mRNA, complete cds.	376	74
1590	gi14043204	Homo sapiens	ribosomal protein, large, P1, clone MGC:15616 IMAGE:3343021, mRNA, complete cds.	376	74
1590	gi190234	Homo sapiens	Human acidic ribosomal phosphoprotein P1 mRNA, complete cds.	376	74
1591	AAW49700	Homo sapiens	Human flavin-containing mono-oxygenase isoform x.	2197	96
1591	gi12006730	Rattus norvegicus	flavin-containing monooxygenase FMO3	2127	72
1591	gi349534	Oryctolagus cuniculus	flavin-containing monooxygenase FMO3	2121	72
1592	AAW49700	Homo sapiens	Human flavin-containing mono-oxygenase isoform x.	2197	96
1592	gi12006730	Rattus norvegicus	flavin-containing monooxygenase FMO3	2127	72
1592	gi349534	Oryctolagus cuniculus	flavin-containing monooxygenase FMO3	2121	72
1593	gi10439603	Homo sapiens	cDNA: FLJ23040 fis, clone LNG02277.	1793	99
1593	gi6633816	Arabidopsis	F1N19.14	141	26

199
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		thaliana			
1593	gi551446	Dictyostelium discoideum	protein kinase	138	22
1594	gi15080438	Homo sapiens	clone MGC:9275 IMAGE:3866562, mRNA, complete cds.	1911	100
1594	gi15082435	Homo sapiens	clone MGC:20227 IMAGE:4558235, mRNA, complete cds.	1911	100
1594	AAB92569	Homo sapiens	Human protein sequence SEQ ID NO:10777.	1911	100
1595	gi10435288	Homo sapiens	cDNA FLJ13315 fis, clone OVARC1001525.	736	100
1595	AAB94572	Homo sapiens	Human protein sequence SEQ ID NO:15358.	736	100
1596	AAG02673	Homo sapiens	Human secreted protein, SEQ ID NO: 6754.	359	100
1596	gi9916	Plasmodium falciparum	liver stage antigen	339	22
1596	gi3549261	Dictyostelium discoideum	interaptin	203	18
1597	gi14993576	Dictyostelium discoideum	deliriumA	218	36
1597	gi6626184	Medicago sativa subsp. x varia	Ran GTPase activating protein	196	29
1597	gi201910	Mus musculus	Tcte-1 peptide	183	32
1598	AAG03347	Homo sapiens	Human secreted protein, SEQ ID NO: 7428.	219	97
1598	AAB28629	Homo sapiens	Human B11Ag1 antigen splice isoform B11C-8.	112	40
1598	AAY82018	Homo sapiens	Human immunogenic prostate tumour protein sequence SEQ ID NO:379.	112	40
1599	AAB08765	Homo sapiens	A human leukocyte and blood related protein (LBAP).	280	56
1599	AAB74718	Homo sapiens	Human membrane associated protein MEMAP-24.	267	60
1599	AAG01306	Homo sapiens	Human secreted protein, SEQ ID NO: 5387.	256	73
1600	gi11385354	Homo sapiens	BAF180 (BAF180) mRNA, complete cds.	8308	100
1600	gi12083875	Homo sapiens	polybromo-1 (PB1) mRNA, complete cds, alternatively spliced.	5129	96
1600	gi951231	Gallus gallus	polybromo 1 protein	4742	89
1601	gi10434904	Homo sapiens	cDNA FLJ13063 fis, clone NT2RP3001678.	2590	99
1601	AAB94448	Homo sapiens	Human protein sequence SEQ ID NO:15085.	2590	99
1601	gi10440369	Homo sapiens	mRNA for FLJ00019 protein, partial cds.	2370	99
1603	AAG01839	Homo sapiens	Human secreted protein, SEQ ID NO: 5920.	446	95
1603	gi459443	Gallus gallus	p104 chicken Rb	57	53
1603	gi2829147	Homo sapiens	lymphocyte-specific protein 1 (LSP1) gene, LSP1-5 allele, partial cds.	57	34

200
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1604	gi1199604	Homo sapiens	Human zinc finger protein C2H2-25 mRNA, complete cds.	2055	100
1604	gi14042415	Homo sapiens	cDNA FLJ14710 fis, clone NT2RP3000632, weakly similar to ZINC FINGER PROTEIN 84.	1574	54
1604	AAB93075	Homo sapiens	Human protein sequence SEQ ID NO:11902.	1574	54
1605	gi22380	Zea mays	CAAT-box DNA binding protein subunit B (NF-YB)	73	34
1605	AAB52040	Homo sapiens	Human secreted protein sequence encoded by gene 29 SEQ ID NO:89.	56	29
1605	gi263304	Bos taurus	insulin-like growth factor-binding protein 4, IGFBP4	71	37
1606	gi14602674	Homo sapiens	clone IMAGE:3929560, mRNA, partial cds.	180	33
1606	gi6537256	Anopheles gambiae	Toll-related protein	104	27
1606	gi466432	Schistocerca americana	Toll protein	87	35
1607	gi4929701	Homo sapiens	CGI-116 protein mRNA, complete cds.	712	99
1607	gi7677064	Homo sapiens	Human protein x 0009 mRNA.	712	99
1607	gi15011990	Homo sapiens	Similar to CGI-116 protein, clone MGC:13336 IMAGE:4272385, mRNA, complete cds.	707	98
1608	gi11640582	Homo sapiens	MSTP037 mRNA, complete cds.	687	100
1608	gi5305335	Mycobacterium tuberculosis	proline-rich mucin homolog	137	40
1608	gi6015454	Hylobates muelleri	dopamine receptor D4	125	37
1609	gi2138290	Homo sapiens	FLII gene, complete cds.	6603	98
1609	gi8698618	Mus musculus	Fliih protein	6300	93
1609	AAB58894	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 602.	4161	96
1610	gi4808631	Homo sapiens	transcription factor-like protein MRGX (MRGX) mRNA, complete cds.	532	94
1610	gi8895210	Homo sapiens	chromosome X MSL3-2 protein mRNA, complete cds.	532	94
1610	gi5931553	Mus musculus	Sid393p	482	85
1611	gi10434431	Homo sapiens	cDNA FLJ12752 fis, clone NT2RP2001174, weakly similar to GASTRULA ZINC FINGER PROTEIN XLCGF46.1.	390	53
1611	AAB94280	Homo sapiens	Human protein sequence SEQ ID NO:14711.	390	53
1611	gi14330448	Homo sapiens	mRNA for zinc finger protein RINZF (RINZF gene).	390	53
1612	gi14714943	Homo sapiens	Similar to outer dense fiber of sperm tails 2, clone MGC:9034 IMAGE:3874501, mRNA, complete cds.	480	26
1612	gi8886477	Rattus norvegicus	cenexin 2	389	25

201
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1612	gi2996006	Homo sapiens	outer dense fiber protein 2/2 mRNA, complete cds.	345	24
1613	gi14714943	Homo sapiens	Similar to outer dense fiber of sperm tails 2, clone MGC:9034 IMAGE:3874501, mRNA, complete cds.	390	24
1613	gi8886477	Rattus norvegicus	cenexin 2	384	23
1613	gi2996000	Rattus norvegicus	outer dense fiber protein ODF84	336	24
1614	gi13543988	Homo sapiens	clone MGC:13047 IMAGE:3626506, mRNA, complete cds.	2572	99
1614	gi10177621	Arabidopsis thaliana	phytoene dehydrogenase-like	1109	49
1614	gi13424781	Caulobacter crescentus	phytoene dehydrogenase-related protein	752	37
1615	gi10435416	Homo sapiens	cDNA FLJ13409 fis, clone PLACE1001716.	2108	100
1615	AAB94636	Homo sapiens	Human protein sequence SEQ ID NO:15515.	2108	100
1615	gi4324457	Schizosaccharomyces pombe	caffeine-induced death protein 1	168	30
1616	AAG04014	Homo sapiens	Human secreted protein, SEQ ID NO: 8095.	284	92
1616	gi9622389	Plasmodium falciparum	variant surface protein	62	53
1616	AAY19771	Homo sapiens	SEQ ID NO 489 from WO9922243.	54	30
1617	gi13774109	Homo sapiens	mitochondria solute carrier protein (MSCP) mRNA, complete cds, alternatively spliced.	151	56
1617	AAB60658	Homo sapiens	Human mitochondrial solute carrier protein hMSC-o.	151	56
1617	gi7578783	Homo sapiens	HT015 protein (HT015) mRNA, complete cds.	151	56
1618	gi12723008	Lactococcus lactis subsp. lactis	UNKNOWN PROTEIN	33	71
1619	gi10436353	Homo sapiens	cDNA FLJ14007 fis, clone Y79AA1002407.	1454	99
1619	AAB94885	Homo sapiens	Human protein sequence SEQ ID NO:16106.	1454	99
1619	AAG01968	Homo sapiens	Human secreted protein, SEQ ID NO: 6049.	647	94
1620	gi10436353	Homo sapiens	cDNA FLJ14007 fis, clone Y79AA1002407.	1335	94
1620	AAB94885	Homo sapiens	Human protein sequence SEQ ID NO:16106.	1335	94
1620	AAG01968	Homo sapiens	Human secreted protein, SEQ ID NO: 6049.	528	83
1621	gi14701866	Dictyostelium discoideum	carmil	307	27
1621	gi1881738	Acanthamoeba castellanii	myosin-I binding protein Acan125	303	27
1621	AAR35072	Homo sapiens	Human t-complex associated testes	119	28

202
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			expressed protein 1.		
1622	gi12053149	Homo sapiens	mRNA; cDNA DKFZp434G2226 (from clone DKFZp434G2226); complete cds.	1178	59
1622	gi1881662	Drosophila melanogaster	kinesin like protein 67a	928	50
1622	gi3136023	Schizosaccharomyces pombe	kinesin-like protein	850	48
1623	gi6069583	Mus musculus	JNK-binding protein JNKBP1	1564	46
1623	gi5911879	Homo sapiens	mRNA; cDNA DKFZp434J046 (from clone DKFZp434J046); partial cds.	1459	88
1623	gi11875643	Myxococcus xanthus	Bap1	182	38
1624	gi12007334	Homo sapiens	IRS-1 PH domain binding protein PHIP mRNA, complete cds.	768	59
1624	gi14286226	Homo sapiens	pleckstrin homology domain interacting protein, clone MGC:15187 IMAGE:3830844, mRNA, complete cds.	768	59
1624	gi12007336	Mus musculus	IRS-1 PH domain binding protein PHIP	768	59
1625	gi11558488	Homo sapiens	mRNA for B-cell lymphoma/leukaemia 11B (BCL11B gene).	4438	100
1625	gi13094147	Mus musculus	zinc finger protein mRit1 beta	4076	93
1625	gi7546793	Mus musculus	CTIP2 protein	3719	86
1626	gi5689375	Rattus norvegicus	tudor repeat associator with PCTAIRE 2	2512	89
1626	gi4586460	Homo sapiens	mRNA for tudor repeat associator with PCTAIRE 2, partial cds.	2498	100
1626	gi13938496	Homo sapiens	clone IMAGE:3834239, mRNA, partial cds.	1431	100
1627	AAG03904	Homo sapiens	Human secreted protein, SEQ ID NO: 7985.	469	100
1627	gi2853287	Homo sapiens	U4/U6-associated RNA splicing factor (PRP3) mRNA, complete cds.	469	100
1627	gi13905000	Mus musculus	RIKEN cDNA 3632413F13 gene	469	100
1628	gi11611734	Homo sapiens	GREB1a (GREB1) mRNA, partial cds, alternatively spliced.	1117	42
1628	gi7264653	Mus musculus	Kiaa0575	953	50
1628	gi11611736	Homo sapiens	GREB1b (GREB1) mRNA, complete cds, alternatively spliced.	913	51
1629	AAAY94673	Homo sapiens	Human zsig83 protein sequence.	948	100
1629	AAAY94674	Homo sapiens	Human zsig83 mature protein sequence.	876	100
1629	AAB60475	Homo sapiens	Human cell cycle and proliferation protein CCYPR-23, SEQ ID NO:23.	197	39
1630	gi14249894	Homo sapiens	clone IMAGE:3502329, mRNA, partial cds.	1321	96
1630	AAB12157	Homo sapiens	Hydrophobic domain protein from clone HP03165 isolated from KB cells.	1321	96
1630	AAAY71116	Homo sapiens	Human Hydrolase protein-14 (HYDRL-14).	1321	96

203
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1631	gi12483904	Rattus norvegicus	zinc finger protein HIT-39	2085	87
1631	gi3417297	Homo sapiens	Human Chromosome 16 BAC clone CIT987SK-A-635H12, complete sequence.	1007	52
1631	gi186774	Homo sapiens	Human Kruppel related zinc finger protein (HTF10) mRNA, complete cds.	746	51
1632	AAB24467	Homo sapiens	Human secreted protein sequence encoded by gene 31 SEQ ID NO:92.	243	34
1632	AAR41899	Homo sapiens	Partial bpTK3 gene prod.	60	34
1632	AAR85926	Homo sapiens	Protein tyrosine-kinase bpTK3 fragment.	60	34
1633	gi13111847	Homo sapiens	COP9 homolog, clone MGC:1297 IMAGE:3503287, mRNA, complete cds.	652	94
1633	gi1730284	Homo sapiens	COP9 signalosome subunit 1 CSN1 (CSN1) mRNA, complete cds.	652	94
1633	AAW68581	Homo sapiens	Human COP9 protein.	652	94
1634	gi10435524	Homo sapiens	cDNA FLJ13491 fis, clone PLACE1004274.	1526	100
1634	AAB94678	Homo sapiens	Human protein sequence SEQ ID NO:15628.	1526	100
1634	gi190074	Homo sapiens	lysyl hydroxylase (PLOD) mRNA, complete cds.	141	31
1635	gi6573115	Mus musculus	p300 transcriptional cofactor JMY	1427	82
1635	gi463250	Mus musculus	Neurofilament protein, high molecular weight subunit (NF-H)	136	30
1635	gi200022	Mus musculus	neurofilament protein	136	30
1636	gi552138	Drosophila melanogaster	tropomyosin isoform 129	140	23
1636	gi3449364	Homo sapiens	intracellular hyaluronic acid binding protein (IHABP) mRNA, complete cds.	148	25
1636	AAW39165	Homo sapiens	Human RHAMM protein.	147	25
1637	gi10834638	Homo sapiens	TBXX T-box containing protein (TBX22) mRNA, complete cds.	2145	100
1637	gi3128382	Mus musculus	MmTbx14	831	49
1637	gi3169261	Mus musculus	T-box transcription factor	824	49
1638	gi11493744	Drosophila melanogaster	STATHMIN-19	124	23
1638	gi11493734	Drosophila melanogaster	stathmin-14	124	23
1638	gi11493732	Drosophila melanogaster	stathmin-1	116	23
1639	gi12005487	Homo sapiens	NPD012 (NPD012) mRNA, complete cds.	246	27
1639	gi10436679	Homo sapiens	cDNA FLJ14262 fis, clone PLACE1001608.	213	27
1639	AAB95811	Homo sapiens	Human protein sequence SEQ ID NO:18804.	213	27
1640	gi11141507	Homo sapiens	beta protein 1 BP1 mRNA, complete cds.	1285	100
1640	AAB72699	Homo sapiens	Human beta-globin gene repressor BP1.	1285	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1640	gi1657869	Mus musculus	Dlx7	879	74
1641	gi10437156	Homo sapiens	cDNA: FLJ21124 fis, clone CAS05964.	664	99
1641	gi15026993	Homo sapiens	partial mRNA for MUC5AC protein (mucin gene, MUC5AC).	66	26
1641	gi4056437	Arabidopsis thaliana	Strong similarity to PFAM PF00069 Eukaryotic protein kinase domain.	83	29
1642	gi8655687	Homo sapiens	mRNA; cDNA DKFZp762E1511 (from clone DKFZp762E1511).	905	100
1642	gi6979930	Homo sapiens	Mam1 mRNA, partial cds.	343	31
1642	gi11127697	Homo sapiens	SYT/SSX4v fusion protein (SSXT/SSX4v fusion) mRNA, partial cds.	164	28
1643	gi14488371	Oryza sativa	casein kinase II alpha subunit, 5'-partial	77	27
1643	gi12697577	Oryza sativa	casein kinase II alpha subunit	77	27
1644	gi3298472	Mus musculus	zinc finger protein	2662	88
1644	gi12653845	Homo sapiens	neighbor of A-kinase anchoring protein 95, clone MGC:1206 IMAGE:3504388, mRNA, complete cds.	361	31
1644	gi6688138	Homo sapiens	mRNA for LA95 protein.	361	31
1645	gi5852425	Homo sapiens	double homeobox protein genes, complete cds.	261	42
1645	gi841206	Homo sapiens	Human D4S2463 homeobox-like gene, partial cds.	256	43
1645	gi1435038	Homo sapiens	Human facioscapulohumeral muscular dystrophy (FSHD) gene region, D4Z4 tandem repeat unit.	254	42
1646	AAE02058	Homo sapiens	Human four disulfide core domain (FDCD)-containing protein.	607	43
1646	gi12655452	Homo sapiens	mRNA for keratin associated protein 4.7 (KRTAP4.7 gene).	552	42
1646	gi200962	Mus musculus	serine 1 ultra high sulfur protein	551	44
1647	gi10281106	Rattus norvegicus	kinesin light chain KLCt	743	90
1647	gi10433849	Homo sapiens	cDNA FLJ12387 fis, clone MAMMA1002637, moderately similar to KINESIN LIGHT CHAIN.	381	55
1647	gi10434570	Homo sapiens	cDNA FLJ12845 fis, clone NT2RP2003307, moderately similar to KINESIN LIGHT CHAIN.	381	55
1648	AAV59712	Homo sapiens	Secreted protein 33-54-1-B9-FL1.	562	98
1648	gi3288566	Oryzias latipes	annexin max1	80	32
1648	gi15075730	Sinorhizobium meliloti	HYPOTHETICAL TRANSMEMBRANE PROTEIN	64	31
1649	gi509241	Homo sapiens	Human mRNA for upstream binding factor (hUBF).	652	45
1649	gi55116	Mus musculus	transcription factor UBF	649	45
1649	gi28971	Homo sapiens	H.sapiens mRNA for autoantigen NOR-90.	638	60
1650	AAB45436	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 110.	75	28
1650	gi14210816	Tupaia	t128	71	37

205
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		herpesvirus			
1650	AAB68057	Homo sapiens	Amino acid sequence of a recombinant human gelatin.	72	29
1651	gi7022392	Homo sapiens	cDNA FLJ10390 fis, clone NT2RM4000104, moderately similar to ZINC FINGER PROTEIN 135.	373	35
1651	AAB92770	Homo sapiens	Human protein sequence SEQ ID NO:11245.	373	35
1651	gi488555	Homo sapiens	Human zinc finger protein ZNF135 mRNA, complete cds.	384	40
1652	gi5817149	Homo sapiens	mRNA; cDNA DKFZp572C163 (from clone DKFZp572C163); partial cds.	3957	100
1652	gi10434142	Homo sapiens	cDNA FLJ12574 fis, clone NT2RM4000996, moderately similar to ZINC FINGER PROTEIN 91.	3217	100
1652	AAB95128	Homo sapiens	Human protein sequence SEQ ID NO:17131.	3217	100
1653	gi4092859	Homo sapiens	p53 regulated PA26-T2 nuclear protein (PA26) mRNA, complete cds.	944	54
1653	gi4092863	Homo sapiens	non-p53 regulated PA26-T1 nuclear protein (PA26) mRNA, complete cds.	942	56
1653	gi13161393	Xenopus laevis	nuclear factor XPA26-T2	907	53
1654	gi6807726	Homo sapiens	mRNA; cDNA DKFZp434K0614 (from clone DKFZp434K0614); partial cds.	3458	100
1654	gi3970852	Homo sapiens	HRIHFB2017 mRNA, partial cds.	1132	100
1654	AAAY86185	Homo sapiens	Nuclear transport protein clone hfb2017 protein sequence.	1132	100
1655	gi11493485	Homo sapiens	PRO2574	384	100
1655	AAW73401	Homo sapiens	Human secreted protein encoded by Gene No. 5.	66	36
1655	gi261363	Mus sp.	T-cell receptor alpha chain, TCR alpha	60	34
1656	gi13359303	Homo sapiens	MAGE-E1a mRNA, complete cds.	2593	99
1656	gi14588553	Homo sapiens	MAGE E1 gene, complete cds, alternative splicing.	2593	99
1656	gi13276667	Homo sapiens	mRNA; cDNA DKFZp761N1924 (from clone DKFZp761N1924); complete cds.	2593	99
1657	gi11493552	Homo sapiens	PRO1933	661	100
1658	gi2555183	Rattus norvegicus	SPA-1 like protein p1294	926	44
1658	gi3970966	Homo sapiens	PAC clone RP5-1140G11 from 14q24.3, complete sequence.	689	50
1658	gi4151328	Homo sapiens	high-risk human papilloma viruses E6 oncoproteins targeted protein E6TP1 alpha mRNA, complete cds.	689	50
1659	gi217346	Drosophila melanogaster	prospero	75	29
1659	gi162792	Bos taurus	alpha-s1-casein precursor	83	28

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1659	gi92	Bos taurus	alpha-S1-casein	83	28
1660	gi1732073	Homo sapiens	Human TBP-associated factor (hTAFII130) mRNA, partial cds.	4736	99
1660	gi2058326	Homo sapiens	H.sapiens mRNA for TAFII135.	4306	85
1660	AAR56494	Homo sapiens	TATA-binding protein-associated factor hTAFII130.	3551	97
1661	gi903599	Homo sapiens	Human Krueppel-type zinc finger protein (ZNF169) gene, final exon, partial cds.	2693	97
1661	gi487783	Homo sapiens	Human zinc finger protein ZNF133.	1373	49
1661	gi9392657	Homo sapiens	PR-domain containing protein 9 (PRDM9) mRNA, complete cds.	1369	51
1662	gi13507379	Homo sapiens	BBS2 (BBS2) mRNA, complete cds.	3715	100
1662	gi14042451	Homo sapiens	cDNA FLJ14729 fis, clone NT2RP3001896.	3684	99
1662	AAB93128	Homo sapiens	Human protein sequence SEQ ID NO:12015.	3684	99
1663	AAV73489	Homo sapiens	Human secreted protein clone yk84_1 protein sequence SEQ ID NO:200.	165	55
1663	gi2674262	murine adenovirus 1	33K protein	72	45
1663	gi1542885	Homo sapiens	H.sapiens PRB4 gene, allele M.	73	48
1664	AAV45305	Homo sapiens	Human secreted protein fragment encoded from gene 6.	235	100
1664	AAW10687	Homo sapiens	Bax omega protein C-terminal peptide used for antibody production.	60	41
1664	gi9947305	Pseudomonas aeruginosa	probable sigma-70 factor, ECF subfamily	62	37
1665	gi453172	Mus musculus	Sax-1	1120	73
1665	gi7248791	Mus musculus	homeobox protein	1120	73
1665	gi7248794	Mus musculus	homeodomain protein	1120	73
1666	gi12053153	Homo sapiens	mRNA; cDNA DKFZp434M1526 (from clone DKFZp434M1526).	5773	100
1666	gi13097273	Homo sapiens	clone IMAGE:3447324, mRNA, partial cds.	1774	100
1666	gi14919417	Mus musculus	Similar to RIKEN cDNA 4930506D01 gene	1277	78
1668	gi7582288	Homo sapiens	BM-008	571	98
1668	gi2992630	Mus musculus	mPRL-2	571	98
1668	gi894159	Homo sapiens	Human protein-tyrosine phosphatase (HU-PP-1) mRNA, partial sequence.	571	98
1669	gi1752736	Saccharomyces cerevisiae	gene required for phosphorylation of oligosaccharides/ has high homology with YJR061w	185	41
1669	AAB35408	Homo sapiens	Human 07CG27 gene protein.	159	30
1669	gi13539605	Paramecium tetraurelia	cyclophilin-RNA interacting protein	170	35
1670	gi7022912	Homo sapiens	cDNA FLJ10713 fis, clone NT2RP3000980.	493	48
1670	AAB93089	Homo sapiens	Human protein sequence SEQ ID NO:11933.	493	48
1670	gi10434247	Homo sapiens	cDNA FLJ12636 fis, clone NT2RM4001905.	168	43

207
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1671	gi11643582	Homo sapiens	PR-domain containing protein 14 (PRDM14) mRNA, complete cds.	3166	100
1671	gi10434076	Homo sapiens	cDNA FLJ12533 fis, clone NT2RM4000202, weakly similar to ZINC FINGER PROTEIN MOK-2.	3166	100
1671	AAB94103	Homo sapiens	Human protein sequence SEQ ID NO:14326.	3166	100
1672	gi7528270	Cucumis sativus	poly(A)-binding protein	69	40
1672	gi4808585	Homo sapiens	KH type splicing regulatory protein (KHSRP) gene, exon 2 and partial cds.	55	55
1672	gi3288485	Homo sapiens	mRNA for chimaeric transcript of collagen type 1 alpha 1 and platelet derived growth factor beta, 202 bp.	53	56
1673	gi14329727	Secale cereale	high molecular weight glutenin subunit y	90	26
1673	gi6684164	Triticum aestivum	glutenin, high molecular weight subunit type y precursor	89	28
1673	AAG03061	Homo sapiens	Human secreted protein, SEQ ID NO: 7142.	44	42
1674	gi10435111	Homo sapiens	cDNA FLJ13196 fis, clone NT2RP3004428, weakly similar to CHROMODOMAIN HELICASE-DNA-BINDING PROTEIN 4.	750	38
1674	AAB94512	Homo sapiens	Human protein sequence SEQ ID NO:15225.	750	38
1674	gi7582284	Homo sapiens	BM-006	646	44
1675	gi14091950	Mus musculus	L-threonine 3-dehydrogenase	333	66
1675	gi10640424	Thermoplasma acidophilum	UDP-glucose 4-epimerase related protein	174	44
1675	gi14324667	Thermoplasma volcanium	NDP-sugar epimerase	174	50
1676	gi12653665	Homo sapiens	clone MGC:2594 IMAGE:3346055, mRNA, complete cds.	536	100
1676	gi1620075	Paramecium bursaria Chlorella virus 1	A403R	69	23
1676	AAG00829	Homo sapiens	Human secreted protein, SEQ ID NO: 4910.	67	36
1677	gi10434750	Homo sapiens	cDNA FLJ12964 fis, clone NT2RP2005732.	2816	100
1677	AAB94389	Homo sapiens	Human protein sequence SEQ ID NO:14949.	2816	100
1677	gi10433565	Homo sapiens	cDNA FLJ12154 fis, clone MAMMA1000468.	2810	99
1678	gi11762184	Arabidopsis thaliana	AT3g07750	508	39
1678	gi5458043	Pyrococcus abyssi	POLYRIBONUCLEOTIDE NUCLEOTIDYLTRANSFERASE related protein	373	34
1678	gi2621767	Methanothermobacter thermautotrophicus	conserved protein	349	33
1679	gi10435659	Homo sapiens	cDNA FLJ13605 fis, clone	793	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			PLACE1010562.		
1679	AAB95488	Homo sapiens	Human protein sequence SEQ ID NO:18023.	793	100
1679	gi7020399	Homo sapiens	cDNA FLJ20356 fis, clone HEP15821.	112	46
1680	gi488555	Homo sapiens	Human zinc finger protein ZNF135 mRNA, complete cds.	1254	57
1680	gi8453103	Homo sapiens	zinc finger protein mRNA, complete cds.	1194	52
1680	gi15012179	Homo sapiens	zinc finger protein 16 (KOX 9), clone MGC:15145 IMAGE:3949487, mRNA, complete cds.	1194	52
1681	gi13991364	Homo sapiens	RING finger protein RIFF mRNA, complete cds.	281	38
1681	gi10437953	Homo sapiens	cDNA: FLJ21786 fis, clone HEP00326.	281	38
1681	AAY56882	Homo sapiens	Human apoptotic inhibitor protein 6 (AIP-6).	281	38
1682	AAY48602	Homo sapiens	Human breast tumour-associated protein 63.	60	36
1682	gi2392814	Mus musculus	PFTAIR kinase	54	28
1682	AAY30087	Homo sapiens	A human cyclin-dependent kinase designated hPFTAIR.	54	28
1683	gi163245	Bos taurus	type II cytokeratin A (no.8)	416	58
1683	gi481	Bos taurus	cytokeratin 8 (370 AA)	416	58
1683	gi12653737	Homo sapiens	keratin 8, clone MGC:1711 IMAGE:3349233, mRNA, complete cds.	456	65
1684	gi11494213	Homo sapiens	Mix.1 homeobox-like protein (MIXL) mRNA, complete cds.	1221	100
1684	gi12698336	Homo sapiens	homeodomain protein MIX (MIX) gene, complete cds.	1221	100
1684	gi4585580	Mus musculus	Mix-like homeobox protein	784	69
1685	gi14249890	Homo sapiens	clone MGC:15766 IMAGE:3501702, mRNA, complete cds.	1442	100
1685	gi10438763	Homo sapiens	cDNA: FLJ22393 fis, clone HRC07880.	1110	74
1685	gi3287375	Mus musculus	C9	689	56
1686	gi10803417	Homo sapiens	mRNA for GPP34-related protein (GPP34R gene).	1244	100
1686	AAB36607	Homo sapiens	Human FLEXHT-29 protein sequence SEQ ID NO:29.	1244	100
1686	gi15082415	Homo sapiens	golgi phosphoprotein 3, clone MGC:20187 IMAGE:4558305, mRNA, complete cds.	877	67
1687	gi10803417	Homo sapiens	mRNA for GPP34-related protein (GPP34R gene).	1459	100
1687	AAB36607	Homo sapiens	Human FLEXHT-29 protein sequence SEQ ID NO:29.	1459	100
1687	gi7022870	Homo sapiens	cDNA FLJ10687 fis, clone NT2RP3000312.	1026	100
1688	gi14804	Bacteriophage phi-80	cI gene (AA 1 - 236)	1236	100

209
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1688	gi476017	Pectobacterium carotovorum	DNA-binding protein	697	59
1688	gi286176	Pseudomonas aeruginosa	negative regulator of pyocin genes	433	40
1689	gi4704279	Schizosaccharomyces pombe	set domain protein; transcriptional silencing	561	46
1689	gi5923931	Homo sapiens	MLL2 protein mRNA, partial cds.	392	53
1689	gi5123787	Homo sapiens	mRNA for trithorax homologue 2.	392	53
1690	AAB90658	Homo sapiens	Human lipocolon (LPC) protein, SEQ ID NO: 201.	639	98
1690	AAB90570	Homo sapiens	Human secreted protein, SEQ ID NO: 108.	568	98
1690	AAG03582	Homo sapiens	Human secreted protein, SEQ ID NO: 7663.	387	97
1691	gi557822	Saccharomyces cerevisiae	mal5, sta1, len: 1367, CAI: 0.3, AMYH_YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	229	24
1691	gi1304387	Saccharomyces cerevisiae var. diastaticus	glucoamylase	229	24
1691	gi545660	Saccharomyces cerevisiae, YNN295, Peptide, 1802 aa	Hkr1p	203	23
1692	gi10437767	Homo sapiens	cDNA: FLJ21628 fis, clone COL08076.	1038	58
1692	gi8099348	Homo sapiens	zinc finger protein (ZFP) mRNA, complete cds.	987	64
1692	gi488555	Homo sapiens	Human zinc finger protein ZNF135 mRNA, complete cds.	994	63
1693	gi13311007	Homo sapiens	protein kinase NYD-SP15 mRNA, complete cds.	845	100
1693	gi13905210	Mus musculus	Similar to protein kinase NYD-SP15	677	93
1693	gi6808241	Homo sapiens	mRNA; cDNA DKFZp434H1720 (from clone DKFZp434H1720); partial cds.	615	100
1694	gi786117	Ensis minor	nuclear protein	242	34
1694	gi5821151	Homo sapiens	mRNA for RNA binding protein, partial cds, clone: R86.	242	35
1694	gi5821153	Homo sapiens	mRNA for RNA binding protein, complete cds.	247	29
1695	gi7619884	Gallus gallus	muscle derived protein	976	52
1695	AAW49042	Homo sapiens	Human low density lipoprotein binding protein LBP-3.	650	59
1695	gi7023731	Homo sapiens	cDNA FLJ11209 fis, clone PLACE1007946.	601	59
1696	gi12653213	Homo sapiens	ribosomal protein L18, clone MGC:8373 IMAGE:2820119, mRNA, complete cds.	309	69
1696	gi337493	Homo sapiens	ribosomal protein L18 (RPL18) mRNA, complete cds.	309	69
1696	gi206724	Rattus norvegicus	ribosomal protein L18	292	60
1697	gi10439510	Homo sapiens	cDNA: FLJ22965 fis, clone KAT10418.	1157	100

210
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1697	AAG01784	Homo sapiens	Human secreted protein, SEQ ID NO: 5865.	471	100
1697	gi2772565	Homo sapiens	Human PAC clone RP3-404F18 from Xq23, complete sequence.	392	100
1698	AAB54170	Homo sapiens	Human pancreatic cancer antigen protein sequence SEQ ID NO:622.	1225	98
1698	gi14042471	Homo sapiens	cDNA FLJ14737 fis, clone NT2RP3002273, weakly similar to SCD6 PROTEIN.	544	50
1698	AAY53005	Homo sapiens	Human secreted protein clone pm749_8 protein sequence SEQ ID NO:16.	544	50
1699	gi2253297	Monodelphis domestica	kinesin homolog	565	80
1699	gi1572665	Dictyostelium discoideum	kinesin-like protein K6	461	46
1699	gi2695866	Mus musculus	kinesin-like protein 2beta	443	42
1700	gi14042307	Homo sapiens	cDNA FLJ14644 fis, clone NT2RP2001756, weakly similar to ZINC FINGER PROTEIN 84.	1265	98
1700	AAB95176	Homo sapiens	Human protein sequence SEQ ID NO:17237.	1265	98
1700	gi15081398	Homo sapiens	kruppel-like zinc finger protein (ZNF300) mRNA, complete cds.	270	44
1701	gi13182773	Homo sapiens	CDA10 mRNA, complete cds.	864	77
1701	gi12006227	Homo sapiens	CDA018 mRNA, complete cds.	864	77
1701	gi13876944	Homo sapiens	NEFA-interacting nuclear protein NIP30 (NIP30) mRNA, complete cds.	864	77
1702	gi12053085	Homo sapiens	mRNA; cDNA DKFZp434A1319 (from clone DKFZp434A1319); complete cds.	1816	100
1702	gi292836	Homo sapiens	Human trichohyalin (TRHY) gene, complete cds.	92	22
1702	AAY30795	Homo sapiens	A human trichohyalin (TRHY) protein.	92	22
1703	AAB30829	Homo sapiens	Amino acid sequence of human signal transduction protein SGT5-1.	1288	99
1703	AAB30830	Homo sapiens	Amino acid sequence of human signal transduction protein SGT5-2.	1288	99
1703	AAB30831	Homo sapiens	Amino acid sequence of human signal transduction protein SGT5-3.	1288	99
1704	gi200964	Mus musculus	serine 2 ultra high sulfur protein	312	50
1704	gi3228237	Homo sapiens	UHS KerB gene.	278	50
1704	gi200962	Mus musculus	serine 1 ultra high sulfur protein	274	47
1705	gi4416181	Mus musculus	ES18	2267	88
1705	gi4416183	Homo sapiens	ES18 mRNA, partial cds.	1109	98
1705	AAB38555	Homo sapiens	Human secreted protein sequence encoded by gene 34 SEQ ID NO:92.	721	90
1706	gi188882	Homo sapiens	Human tracheo-bronchial mucin (MUC4) mRNA, partial cds.	2181	74
1706	gi454154	Homo sapiens	intestinal mucin (MUC2) mRNA, complete cds.	1490	29
1706	gi14973269	Streptococcus pneumoniae	cell wall surface anchor family protein	1283	23

211
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1707	gi12667440	Homo sapiens	NIR1 mRNA, complete cds.	5069	99
1707	AAW59357	Homo sapiens	Human retinal degeneration B1 polypeptide (hrdgB1).	5069	99
1707	gi9367838	Homo sapiens	mRNA full length insert cDNA clone EUROIMAGE 2004661.	4731	99
1708	gi387055	Cricetulus griseus	RNA polymerase II largest subunit	628	40
1708	gi825713	Homo sapiens	H.sapiens gene for RNA pol II largest subunit, exon 1.	631	38
1708	gi36124	Homo sapiens	H.sapiens mRNA for RNA polymerase II largest subunit.	631	38
1709	gi2618752	Takifugu rubripes	zinc finger protein	1149	67
1709	gi1020145	Homo sapiens	Human DNA binding protein (HPF2) mRNA, complete cds.	605	37
1709	gi2843171	Homo sapiens	zinc finger protein mRNA, complete cds.	605	38
1710	gi2618752	Takifugu rubripes	zinc finger protein	1496	66
1710	gi10434850	Homo sapiens	cDNA FLJ13029 fis, clone NT2RP3001057, moderately similar to ZINC FINGER PROTEIN 91.	707	40
1710	AAB95278	Homo sapiens	Human protein sequence SEQ ID NO:17486.	707	40
1711	gi7328175	Homo sapiens	mRNA; cDNA DKFZp762H157 (from clone DKFZp762H157); complete cds.	2746	100
1711	AAB53356	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:896.	2746	100
1711	gi31283	Homo sapiens	Human mRNA for ezrin.	2743	99
1712	gi7328175	Homo sapiens	mRNA; cDNA DKFZp762H157 (from clone DKFZp762H157); complete cds.	2404	98
1712	AAB53356	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:896.	2404	98
1712	gi31283	Homo sapiens	Human mRNA for ezrin.	2401	98
1713	gi7328175	Homo sapiens	mRNA; cDNA DKFZp762H157 (from clone DKFZp762H157); complete cds.	2589	99
1713	AAB53356	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:896.	2589	99
1713	gi31283	Homo sapiens	Human mRNA for ezrin.	2586	99
1714	gi9963808	Homo sapiens	sentrin/SUMO-specific protease (SEN7) mRNA, complete cds.	1253	100
1714	gi13276699	Homo sapiens	mRNA; cDNA DKFZp564G1816 (from clone DKFZp564G1816); complete cds.	1170	98
1714	AAB94994	Homo sapiens	Human protein sequence SEQ ID NO:16621.	379	55
1715	gi488555	Homo sapiens	Human zinc finger protein ZNF135 mRNA, complete cds.	1258	60
1715	gi5441615	Canis familiaris	zinc finger protein	1255	59
1715	gi10437767	Homo sapiens	cDNA: FLJ21628 fis, clone COL08076.	1248	65
1716	gi10438660	Homo sapiens	cDNA: FLJ22321 fis, clone	1878	100

212
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			HRC05390.		
1716	gi12655057	Homo sapiens	clone MGC:2574 IMAGE:3051044, mRNA, complete cds.	1878	100
1716	AAE03642	Homo sapiens	Human extracellular matrix and cell adhesion molecule-6 (XMAD-6).	1878	100
1717	gi4557143	Rattus norvegicus	zinc finger protein RIN ZF	320	33
1717	gi14330448	Homo sapiens	mRNA for zinc finger protein RINZF (RINZF gene).	320	32
1717	gi726284	Mus musculus	polyomavirus late initiator promoter binding protein	333	28
1718	gi8272464	Homo sapiens	human endogenous retrovirus W gagC3.37 G gag (gag) gene, complete cds.	318	55
1718	AAB07702	Homo sapiens	Protein encoded by the endogenetic fragment of HERV-W.	318	55
1718	gi5726236	multiple sclerosis associated retrovirus element	gag polyprotein	306	54
1719	gi13623489	Homo sapiens	clone IMAGE:4109498, mRNA, partial cds.	1007	99
1719	gi7209311	Homo sapiens	mRNA for FLJ00005 protein, partial cds.	510	100
1719	gi683545	Chironomus pallidivittatus	gamma protein constant region	85	34
1720	gi10433129	Homo sapiens	cDNA FLJ11792 fis, clone HEMBA1006121.	849	100
1720	AAB93829	Homo sapiens	Human protein sequence SEQ ID NO:13646.	849	100
1720	gi10440076	Homo sapiens	cDNA: FLJ23401 fis, clone HEP18813.	259	42
1721	gi10438872	Homo sapiens	cDNA: FLJ22471 fis, clone HRC10529.	1681	100
1721	AAAY86509	Homo sapiens	Human gene 70-encoded protein fragment, SEQ ID NO:424.	696	100
1721	AAAY86510	Homo sapiens	Human gene 70-encoded protein fragment, SEQ ID NO:425.	436	100
1722	gi14039857	Homo sapiens	testes development-related NYD-SP22 mRNA, complete cds.	275	92
1722	gi10444372	Phaseolus coccineus	Kunitz trypsin inhibitor protein	60	38
1722	gi506139	Zea mays	Ec metallothionein class II protein	58	27
1723	gi31680	Homo sapiens	Human mRNA for alpha-glucocorticoid receptor (clone OB7).	2027	100
1723	gi31682	Homo sapiens	Human mRNA for beta-glucocorticoid receptor (clone OB10).	2027	100
1723	gi2218074	Homo sapiens	glucocorticoid receptor (GRL) gene, intron H exon 9alpha, and complete cds.	2027	100
1724	gi13529542	Mus musculus	RIKEN cDNA 4731402F03 gene	1525	96
1724	gi3777529	Homo sapiens	retinoic acid receptor responder 3 (RARRES3) mRNA, complete cds.	69	41

213
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1724	AAW60727	Homo sapiens	Tazarotene inducible Gene-3 (TIG3) protein.	69	41
1725	gi13436404	Homo sapiens	clone MGC:3855 IMAGE:2905681, mRNA, complete cds.	1336	100
1725	gi32022	Homo sapiens	H.sapiens HAP1 gene for AP endonuclease 1.	1336	100
1725	gi178747	Homo sapiens	Human apurinic/apyrimidinic endonuclease (HAP1h) mRNA, complete cds.	1336	100
1726	gi10440181	Homo sapiens	cDNA: FLJ23477 fis, clone HSI15732.	814	100
1726	gi13383233	Mus musculus	bHLH factor Hes7	102	32
1726	gi211358	Gallus gallus	alpha-1 collagen type IX	66	55
1727	AAG03366	Homo sapiens	Human secreted protein, SEQ ID NO: 7447.	355	97
1727	gi1504095	Mus musculus	DNA-binding protein	85	50
1727	gi1666910	Mus musculus	neurogenin 2	85	50
1728	gi12583669	Homo sapiens	mRNA for bHLH protein DEC2, complete cds.	2509	100
1728	AAB70692	Homo sapiens	Human DEC2a protein sequence SEQ ID NO:2.	2509	100
1728	AAB70693	Homo sapiens	Human DEC2b protein sequence SEQ ID NO:12.	2498	99
1729	gi11320940	Homo sapiens	SCAND2 mRNA, complete cds.	1607	99
1729	gi11345048	Homo sapiens	SCAN domain-containing protein 2 (SCAND2) gene, complete cds, alternatively spliced.	1607	99
1729	gi10434473	Homo sapiens	cDNA FLJ12782 fis, clone NT2RP2001869, moderately similar to ZINC FINGER PROTEIN 191.	708	100
1730	gi10437365	Homo sapiens	cDNA: FLJ21290 fis, clone COL01954.	813	100
1730	gi2447166	Paramecium bursaria Chlorella virus 1	a611R	64	30
1730	gi8895105	Pseudomonas phage D3	Cro	60	47
1731	gi10440052	Homo sapiens	cDNA: FLJ23384 fis, clone HEP16468.	1828	99
1731	gi9947369	Pseudomonas aeruginosa	agmatinase	1014	61
1731	gi7544034	Streptomyces coelicolor A3(2)	agmatinase (fragment)	646	42
1732	gi14575683	synthetic construct	glutamate dehydrogenase	1453	98
1732	gi31707	Homo sapiens	Human mRNA for glutamate dehydrogenase (EC 1.4.1.3., GDH).	1453	98
1732	gi31799	Homo sapiens	Human mRNA for glutamate dehydrogenase (Glud-1, EC 1.4.1.3).	1453	98
1733	gi13905148	Mus musculus	RIKEN cDNA 5430417M23 gene	1400	83
1733	gi439146	Saccharomyces cerevisiae	B-type cyclin	96	32
1733	gi433702	Schizosaccharomyces pombe	mitotic B-type cyclin	95	23
1734	gi13936547	Homo sapiens	formin-binding protein 17 (FBP17)	3194	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			mRNA, partial cds.		
1734	gi10435680	Homo sapiens	cDNA FLJ13619 fis, clone PLACE1010926, weakly similar to HYPOTHETICAL 72.2 KD PROTEIN C12C2.05C IN CHROMOSOME II.	1680	100
1734	AAB94690	Homo sapiens	Human protein sequence SEQ ID NO:15657.	1680	100
1735	gi6164743	Homo sapiens	F-box protein Fbx20 (FBX20) mRNA, partial cds.	1002	100
1735	AAY83093	Homo sapiens	F-box protein FBP-25.	1002	100
1735	AAY83072	Homo sapiens	F-box motif of FBP-25.	261	100
1736	gi1699293	human, megakaryoblastoid cell line MOLM-1, chronic myelocytic leukemia patient, mRNA Partial Mutant, 916 nt]. [Homo sapiens	Evi-1=Evi-1 protein {3' region, deletion region}	57	40
1736	gi15160392	Agrobacterium tumefaciens	AGR_L_3548p	56	33
1736	gi2642488	Sus scrofa	cytochrome b5	56	42
1737	gi14530677	Homo sapiens	mRNA for WNT14, complete cds.	1611	98
1737	gi2623871	Gallus gallus	Wnt-14 protein	1406	85
1737	gi2605811	Homo sapiens	Wnt-like protein Wnt14 gene, partial cds.	681	100
1738	gi14530677	Homo sapiens	mRNA for WNT14, complete cds.	1807	97
1738	gi2623871	Gallus gallus	Wnt-14 protein	1554	82
1738	gi2605811	Homo sapiens	Wnt-like protein Wnt14 gene, partial cds.	681	100
1739	gi12805039	Homo sapiens	clone IMAGE:3453830, mRNA, partial cds.	2504	100
1739	gi13276673	Homo sapiens	mRNA; cDNA DKFZp761A132 (from clone DKFZp761A132); complete cds.	1143	100
1739	gi2290392	Strongyloides stercoralis	IgG and IgE immunoreactive antigen recognized by sera from patients with strongyloidiasis	85	50
1740	gi10435475	Homo sapiens	cDNA FLJ13456 fis, clone PLACE1003258, weakly similar to EARLY EMBRYOGENESIS ZYG-11 PROTEIN.	886	98
1740	AAB94660	Homo sapiens	Human protein sequence SEQ ID NO:15579.	886	98
1740	gi2769562	Homo sapiens	H.sapiens mRNA for ZYG homologue.	623	30
1741	gi624922	Samia cynthia	arylphorin (partial)	71	23
1742	AAG00335	Homo sapiens	Human secreted protein, SEQ ID NO: 4416.	863	99
1742	gi5668737	Mus musculus	UBE-1c2	752	53
1742	gi14042700	Homo sapiens	cDNA FLJ14868 fis, clone PLACE1002395, weakly similar to	448	56

215
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			Mus musculus mRNA for UBE-1c1, UBE-1c2, UBE-1c3.		
1743	gi14043850	Homo sapiens	Similar to splicing factor (45kD), clone MGC:14439 IMAGE:4303675, mRNA, complete cds.	1470	99
1743	AAB08622	Homo sapiens	Amino acid sequence of a human DRT111 polypeptide.	1470	99
1743	gi3746840	Homo sapiens	45kDa splicing factor mRNA, complete cds.	1460	98
1744	gi14020951	Homo sapiens	mRNA for FLJ00029 protein, partial cds.	1293	99
1744	gi7019911	Homo sapiens	cDNA FLJ20059 fis, clone COL01349.	979	48
1744	gi7243777	Drosophila melanogaster	Diablo	749	34
1745	AAG02849	Homo sapiens	Human secreted protein, SEQ ID NO: 6930.	301	98
1745	gi4959731	Tetrahymena pyriformis	chaperonin-containing-TCP1 theta subunit	89	22
1745	gi940781	Methanopyrus kandleri	thermosome, chaperonin	104	25
1746	gi10439289	Homo sapiens	cDNA: FLJ22774 fis, clone KAIA1575.	2313	100
1746	AAB95753	Homo sapiens	Human protein sequence SEQ ID NO:18665.	666	48
1746	AAE01232	Homo sapiens	Human gene 1 encoded secreted protein HMIAJ30, SEQ ID NO:94.	1013	47
1747	gi10944334	Homo sapiens	Hu-Claspin mRNA, complete cds.	6812	99
1747	gi10944336	Xenopus laevis	Claspin	1680	54
1747	AAW03626	Homo sapiens	Human thyrotropin GPR N-terminal sequence.	212	37
1748	gi338973	Homo sapiens	Human T-cell receptor active beta-chain (V6-D-J-C) mRNA, clone PL5.10.	60	39
1749	gi12328443	Homo sapiens	mRNA for PAPA-1, complete cds.	1810	99
1749	gi12328441	Mus musculus	PAPA-1	1695	93
1749	AAG01201	Homo sapiens	Human secreted protein, SEQ ID NO: 5282.	623	100
1750	gi11762214	Arabidopsis thaliana	AT3g04520	242	41
1750	gi2654615	Pseudomonas aeruginosa	L-allo-threonine aldolase	242	45
1750	gi4982322	Thermotoga maritima	L-allo-threonine aldolase	239	40
1751	gi1854550	Mus musculus	red-1	607	96
1751	gi13435627	Mus musculus	Similar to nucleoredoxin	607	96
1751	gi4056568	Zea mays	PDI-like protein	116	52
1752	gi10440418	Homo sapiens	mRNA for FLJ00044 protein, partial cds.	2034	66
1752	gi10433645	Homo sapiens	cDNA FLJ12221 fis, clone MAMMA1001091.	1086	65
1752	AAB93931	Homo sapiens	Human protein sequence SEQ ID NO:13927.	1086	65
1753	gi13544086	Homo sapiens	clone MGC:13251	950	100

216
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			IMAGE:4053416, mRNA, complete cds.		
1753	gi4101720	Mus musculus	lymphocyte specific formin related protein	209	34
1753	gi6708478	Mus musculus	formin-like protein	209	34
1754	gi5531811	Homo sapiens	BUP	608	100
1754	gi6563294	Homo sapiens	PIL protein mRNA, complete cds.	608	100
1754	AAW71684	Homo sapiens	Amino acid sequence of the human tumorigenesis associated protein.	608	100
1755	gi10799166	Homo sapiens	protein kinase Njmu-R1 mRNA, complete cds.	1884	99
1755	gi14290030	Human immunodeficiency virus type 1	pol protein	68	30
1756	gi10440275	Homo sapiens	cDNA: FLJ23550 fis, clone LNG08970.	136	22
1756	gi7021367	Drosophila melanogaster	c11.1	132	19
1757	gi7021367	Drosophila melanogaster	c11.1	117	19
1758	gi14547148	Homo sapiens	mRNA for EGLN2 protein.	2183	100
1758	gi13649965	Mus musculus	cell growth regulator Falkor	1918	87
1758	gi14547241	Mus musculus	EGLN2 protein	1868	86
1759	gi13810677	Mus musculus	dachshund-like protein DACH2	737	88
1759	gi6651521	Gallus gallus	Dach2 protein	622	77
1759	gi5102584	Homo sapiens	mRNA full length insert cDNA clone EUROIMAGE 381801.	443	82
1760	AAB74314	Homo sapiens	Human splicing factor 2 protein.	722	100
1760	gi7022544	Homo sapiens	cDNA FLJ10482 fis, clone NT2RP2000153, weakly similar to GAR2 PROTEIN.	616	99
1760	AAB92868	Homo sapiens	Human protein sequence SEQ ID NO:11452.	616	99
1761	gi13345407	Mus musculus	radial spokehead-L protein	1841	53
1761	gi12053039	Homo sapiens	mRNA; cDNA DKFZp434I0515 (from clone DKFZp434I0515); complete cds.	1787	62
1761	gi2905895	Strongylocentrotus purpuratus	radial spokehead	1703	61
1762	gi2529737	Xenopus laevis	ER1	979	44
1762	AAAY28859	Homo sapiens	Human mesoderm induction early response protein ER1.	965	52
1762	AAAY15835	Homo sapiens	A human er1 protein.	965	52
1763	gi10435949	Homo sapiens	cDNA FLJ13815 fis, clone THYRO1000381.	2949	99
1763	AAB95581	Homo sapiens	Human protein sequence SEQ ID NO:18241.	2949	99
1763	gi861295	Caenorhabditis elegans	similar to S. cerevisiae HAP4 transcriptional activator (SP:HAP4_YEAST, P14064); similar to regulatory domain of PI3-kinase P85-alpha subunit (phosphatidylinositol 3-kinase) and BCR, the product of the breakpoint cluster region gene	189	27

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1764	gi5911171	Homo sapiens	mucin 11 (MUC11) mRNA, partial cds.	4817	99
1764	AAY59288	Homo sapiens	Human MUC11 polypeptide.	4817	99
1764	gi1834503	Homo sapiens	H.sapiens MUC5B gene (partial).	621	29
1765	gi5911881	Homo sapiens	mRNA; cDNA DKFZp434K156 (from clone DKFZp434K156); partial cds.	2384	100
1765	gi5830706	variola minor virus	A42R protein	64	32
1765	gi516427	Variola virus	A37R	63	32
1766	gi7328104	Homo sapiens	mRNA; cDNA DKFZp761N1024 (from clone DKFZp761N1024); partial cds.	3570	99
1766	gi10440043	Homo sapiens	cDNA: FLJ23377 fis, clone HEP16237.	2874	99
1766	gi4884314	Homo sapiens	mRNA; cDNA DKFZp566I0746 (from clone DKFZp566I0746); partial cds.	852	89
1767	gi10435911	Homo sapiens	cDNA FLJ13787 fis, clone PLACE4000670.	491	51
1767	AAB94757	Homo sapiens	Human protein sequence SEQ ID NO:15817.	491	51
1767	gi6176338	Homo sapiens	ubiquitous tetratricopeptide containing protein RoXaN mRNA, partial cds.	488	51
1768	gi1620100	Paramecium bursaria Chlorella virus 1	Pro- and Glu-rich, PENPEV (10x); similar to Streptococcus B antigen, corresponds to Swiss-Prot Accession Number P27951	112	33
1768	gi14915682	Streptococcus pyogenes	fibronectin-binding protein Fba	116	23
1768	gi15149010	Homo sapiens	variably charged X-C (VCXC) mRNA, complete cds.	109	29
1769	gi15147877	Mus musculus	Spred-2	1170	92
1769	gi15147875	Mus musculus	Spred-1	475	56
1769	AAB45199	Homo sapiens	Human secreted protein sequence encoded by gene 28 SEQ ID NO:140.	251	62
1770	gi442368	Rattus norvegicus	neuronal olfactomedin-related ER localized protein	371	33
1770	gi14250608	Homo sapiens	Similar to olfactomedin related ER localized protein, clone MGC:1341 IMAGE:3349741, mRNA, complete cds.	371	34
1770	gi15079891	Homo sapiens	clone MGC:19671 IMAGE:3352603, mRNA, complete cds.	371	34
1771	gi9930614	Homo sapiens	steroid receptor RNA activator isoform 3 mRNA, complete cds.	1241	100
1771	gi9930610	Homo sapiens	steroid receptor RNA activator isoform 1 mRNA, complete cds.	1224	99
1771	gi9930612	Homo sapiens	steroid receptor RNA activator isoform 2 mRNA, complete cds.	1218	98
1772	gi12053105	Homo sapiens	mRNA; cDNA DKFZp434K111 (from clone DKFZp434K111);	539	32

218
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			complete cds.		
1772	gi7019911	Homo sapiens	cDNA FLJ20059 fis, clone COL01349.	473	31
1772	gi2282582	Mus musculus	actin-binding protein	466	29
1773	gi10437750	Homo sapiens	cDNA: FLJ21616 fis, clone COL07477.	375	100
1773	gi13938370	Homo sapiens	microtubule-associated protein, RP/EB family, member 2, clone MGC:1279 IMAGE:2989643, mRNA, complete cds.	74	40
1773	gi1292868	Homo sapiens	H.sapiens mRNA for novel T-cell activation protein.	74	40
1774	gi12053075	Homo sapiens	mRNA; cDNA DKFZp434N1817 (from clone DKFZp434N1817); complete cds.	4044	99
1774	gi6760043	Homo sapiens	variable charge protein on X with eight repeats (VCX-8r) mRNA, complete cds.	69	33
1774	gi11934656	Homo sapiens	variably charged X-A (VCXA) mRNA, complete cds.	69	33
1775	AAB26846	Homo sapiens	Human MASL1 protein sequence.	248	33
1775	gi3293318	Caenorhabditis elegans	leucine-rich repeat protein SOC-2	243	29
1775	gi3252977	Caenorhabditis elegans	Ras-binding protein SUR-8	243	29
1776	gi7650364	Rattus norvegicus	delta Kalirin-7	162	23
1776	gi3108193	Rattus norvegicus	Duo	159	22
1776	gi7767545	Rattus norvegicus	Kalirin-7c isoform	162	23
1777	gi5802824	Homo sapiens	endogenous retrovirus HERV-K109, complete sequence.	841	64
1777	gi5802810	Homo sapiens	endogenous retrovirus HERV-K101, complete sequence.	827	62
1777	gi5802814	Homo sapiens	endogenous retrovirus HERV-K103, complete sequence.	839	63
1778	gi11993915	Homo sapiens	HOXB9 gene, exon 2 and complete cds.	1334	100
1778	gi440955	Mus sp.	Abdominal-B homolog=HoxB9	1285	96
1778	gi4322080	Danio rerio	homeobox protein	897	70
1779	gi11385660	Homo sapiens	CTCL tumor antigen se57-1 mRNA, complete cds.	1654	99
1779	gi4103950	Sus scrofa	tuftelin	181	27
1779	gi3986746	Bos taurus	tuftelin	184	27
1780	AAB19406	Homo sapiens	Amino acid sequence of a human secreted protein.	683	75
1780	AAB47275	Homo sapiens	hOAT4.	663	74
1780	AAV52386	Homo sapiens	Human transmembrane protein HP02000.	630	68
1781	gi3139158	Homo sapiens	LINE-1 like protein mRNA, partial cds.	318	100
1781	AAB44391	Homo sapiens	Sequence homologous to human secreted protein encoded by gene 8.	153	49
1781	gi5070622	Homo sapiens	retrotransposon L1 insertion in X-	159	46

219
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			linked retinitis pigmentosa locus, complete sequence.		
1782	gi6841266	Homo sapiens	HSPC308	206	58
1782	AAB57128	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1706.	202	58
1782	gi7020993	Homo sapiens	cDNA FLJ20719 fis, clone HEP17004.	202	58
1783	AAG01500	Homo sapiens	Human secreted protein, SEQ ID NO: 5581.	515	99
1783	gi9799628	Drosophila melanogaster	Nut2	408	60
1783	gi3879644	Caenorhabditis elegans	T09A5.6	282	44
1784	gi14043273	Homo sapiens	clone IMAGE:3505029, mRNA, partial cds.	84	26
1784	gi6503210	Cervus nippon	amelogenin	65	43
1784	AAB59012	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 720.	84	26
1785	gi10580910	Halobacterium sp. NRC-1	Vng1407c	83	48
1785	gi2635307	Bacillus subtilis	ysmA	72	28
1785	gi15024997	Clostridium acetobutylicum	Predicted esterase	68	45
1786	gi13539682	Homo sapiens	golgi-associated microtubule-binding protein HOOK3 mRNA, complete cds.	2320	100
1786	gi3005085	Homo sapiens	hook1 protein (HOOK1) mRNA, complete cds.	1425	61
1786	gi15079605	Homo sapiens	Similar to hook1 protein, clone MGC:10642 IMAGE:3959931, mRNA, complete cds.	1425	61
1787	gi7018517	Homo sapiens	mRNA; cDNA DKFZp434M035 (from clone DKFZp434M035).	1379	99
1787	gi8118231	Homo sapiens	B17 mRNA, complete cds.	836	100
1787	gi8118229	Homo sapiens	B17 long form mRNA, complete cds.	822	99
1788	gi1899230	Homo sapiens	Human iroquois-class homeodomain protein IRX-2a mRNA, partial cds.	206	33
1788	gi9965418	Mus musculus	iroquois-class homeobox protein IRX2	212	46
1788	gi7576708	Mus musculus	iroquois-class homeobox protein Irx6	212	46
1789	AAG02679	Homo sapiens	Human secreted protein, SEQ ID NO: 6760.	297	100
1789	gi295671	Saccharomyces cerevisiae	selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III	123	24
1789	gi531492	Saccharomyces cerevisiae	RRP3p	123	27
1790	gi12053347	Homo sapiens	mRNA; cDNA DKFZp586M1120 (from clone DKFZp586M1120); complete cds.	1154	100
1790	gi12655852	Mus musculus	protein phosphatase-1 regulatory	236	41

220
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			subunit 7		
1790	gi4633066	Homo sapiens	protein phosphatase-1 regulatory subunit 7 (PPP1R7) gene, exon 11, complete cds and alternatively spliced product.	234	42
1791	gi13435972	Homo sapiens	clone IMAGE:3450558, mRNA, partial cds.	862	94
1791	gi600118	Zea mays	extensin-like protein	422	32
1791	gi8163634	Streptococcus pneumoniae	surface protein PspC	396	31
1792	gi1163177	Homo sapiens	Human inducible poly(A)-binding protein mRNA, complete cds.	1447	75
1792	gi2801403	Homo sapiens	polyadenylate binding protein mRNA, complete cds.	1447	75
1792	gi13096978	Mus musculus	Similar to poly(A)-binding protein, cytoplasmic 4 (inducible form)	1440	75
1793	gi13276231	Homo sapiens	mRNA for FYVE and coiled-coil domain containing 1 (FYCO1 gene).	1881	99
1793	gi10438562	Homo sapiens	cDNA: FLJ22251 fis, clone HRC02686.	161	36
1793	gi7023688	Homo sapiens	cDNA FLJ11183 fis, clone PLACE1007488, weakly similar to PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR.	164	42
1794	gi4456148	Gallus gallus	paralemmin	223	44
1794	gi4456677	Mus musculus	paralemmin	355	39
1794	gi4456613	Homo sapiens	mRNA for paralemmin.	345	38
1795	gi600118	Zea mays	extensin-like protein	245	27
1795	gi9280319	Arabidopsis thaliana	extensin protein-like	219	25
1795	gi1572721	Homo sapiens	Human megakaryocyte stimulating factor mRNA, complete cds.	218	22
1796	gi10439695	Homo sapiens	cDNA: FLJ23114 fis, clone LNG07880.	749	100
1796	gi942602	Ancylostoma caninum	anticoagulant protein 6 precursor	68	28
1796	AAB92768	Homo sapiens	Human protein sequence SEQ ID NO:11240.	74	32
1797	gi14042627	Homo sapiens	cDNA FLJ14825 fis, clone OVARC1000781.	1978	100
1797	AAB94534	Homo sapiens	Human protein sequence SEQ ID NO:15270.	1978	100
1797	gi3319457	Caenorhabditis elegans	contains similarity to O-linked GlcNAc transferases	94	27
1798	gi6166507	Mus musculus	RP42	428	43
1798	gi9896486	Homo sapiens	RP42 protein mRNA, complete cds.	428	43
1798	gi14550461	Homo sapiens	RP42 homolog, clone MGC:15099 IMAGE:3939758, mRNA, complete cds.	428	43
1799	gi13278936	Homo sapiens	Similar to RIKEN cDNA 5430432M24 gene, clone MGC:4675 IMAGE:3532660, mRNA, complete cds.	1527	100
1799	gi1710282	Homo sapiens	Human clone 23803 mRNA, partial	337	48

221
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			cds.		
1799	gi19923	Nicotiana tabacum	pistil extensin like protein, partial CDS	117	31
1801	gi12652727	Homo sapiens	clone IMAGE:3352566, mRNA, partial cds.	2142	100
1801	AAB95536	Homo sapiens	Human protein sequence SEQ ID NO:18138.	2116	99
1801	gi1020145	Homo sapiens	Human DNA binding protein (HPF2) mRNA, complete cds.	863	47
1802	gi12653227	Homo sapiens	interleukin enhancer binding factor 2, 45kD, clone MGC:8391 IMAGE:2820505, mRNA, complete cds.	1050	100
1802	gi532313	Homo sapiens	Human nuclear factor NF45 mRNA, complete cds.	1050	100
1802	AAB58802	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 510.	1050	100
1803	gi10644754	Caenorhabditis elegans	GATA-type transcription factor	92	32
1803	gi3880139	Caenorhabditis elegans	Similarity to Yeast nitrogen regulatory protein GLN3 (PIR Acc. No. S22280), contains similarity to Pfam domain: PF00320 (GATA zinc finger), Score=20.9, E-value=0.0044, N=1	92	32
1803	gi5706504	Schizosaccharom yces pombe	Highly conserved ring finger; similar to yeast apc11 SPC component	74	35
1804	gi14043223	Homo sapiens	clone MGC:15677 IMAGE:3350001, mRNA, complete cds.	888	100
1804	gi14326566	Arabidopsis thaliana	AT4g20350/F9F13_6	177	36
1804	gi15155438	Agrobacterium tumefaciens	AGR_C_894p	115	25
1805	gi14043223	Homo sapiens	clone MGC:15677 IMAGE:3350001, mRNA, complete cds.	244	93
1805	gi15155438	Agrobacterium tumefaciens	AGR_C_894p	91	26
1805	gi3372624	Mus musculus	EGF-like growth factor receptor ErbB4 intracellular domain	65	27
1806	gi14043223	Homo sapiens	clone MGC:15677 IMAGE:3350001, mRNA, complete cds.	231	100
1806	gi15155438	Agrobacterium tumefaciens	AGR_C_894p	84	26
1806	gi1592321	Methanococcus jannaschii	GMP synthase (guaA)	84	33
1807	gi2407913	Homo sapiens	H.sapiens MLN50 mRNA.	809	91
1807	AAW25767	Homo sapiens	Human Lasp-1.	809	91
1807	gi13506795	Rattus norvegicus	LASP-1	758	85
1808	AAB65641	Homo sapiens	Novel protein kinase, SEQ ID NO:	1259	90

222

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			168.		
1808	gi2645810	Mus musculus	Pftaire-1	1195	68
1808	AAY30087	Homo sapiens	A human cyclin-dependent kinase designated hPFTAIRe.	1194	68
1809	gi10435062	Homo sapiens	cDNA FLJ13163 fis, clone NT2RP3003656.	2604	99
1809	AAB95343	Homo sapiens	Human protein sequence SEQ ID NO:17625.	2604	99
1809	gi10436604	Homo sapiens	cDNA FLJ14205 fis, clone NT2RP3003155.	2273	100
1810	AAG01851	Homo sapiens	Human secreted protein, SEQ ID NO: 5932.	445	98
1810	gi483843	Bos taurus	50 kDa protein	88	35
1810	AAB56241	Homo sapiens	Human secreted protein sequence encoded by gene 8 SEQ ID NO:335.	42	50
1811	gi13097573	Homo sapiens	Similar to thiosulfate sulfurtransferase (rhodanese), clone MGC:10492 IMAGE:3611253, mRNA, complete cds.	1598	100
1811	gi432376	Homo sapiens	Human rohu mRNA for rhodanese.	1575	98
1811	gi1741864	Rattus norvegicus	mercaptopyruvate sulfurtransferase	1385	85
1812	AAY00294	Homo sapiens	Human secreted protein encoded by gene 37.	240	100
1812	gi3366918	Arabidopsis thaliana	NADH dehydrogenase subunit F	76	31
1812	gi5881743	Arabidopsis thaliana	NADH dehydrogenase ND5	76	31
1813	gi1262852	Mus musculus	M17 protein	134	29
1813	gi486	Bos taurus	epidermal keratin VII	63	30
1813	gi3287141	Human immunodeficiency virus type 1	vpu	67	37
1814	gi13539682	Homo sapiens	golgi-associated microtubule-binding protein HOOK3 mRNA, complete cds.	193	32
1814	gi3005087	Homo sapiens	hook2 protein (HOOK2) mRNA, complete cds.	175	33
1814	AAY82321	Homo sapiens	Human protein transport molecule (PTAM) SEQ ID NO:5.	175	33
1815	gi10434781	Homo sapiens	cDNA FLJ12985 fis, clone NT2RP3000050, moderately similar to ZINC FINGER PROTEIN 91.	582	100
1815	AAB94396	Homo sapiens	Human protein sequence SEQ ID NO:14964.	582	100
1815	gi7023216	Homo sapiens	cDNA FLJ10891 fis, clone NT2RP4002078, weakly similar to ZINC FINGER PROTEIN 91.	483	55
1816	gi915208	Sus scrofa	gastric mucin	339	29
1816	gi7332056	Caenorhabditis elegans	contains similarity to Pfam family PF00078 (Reverse transcriptase (RNA-dependent)), score=79.6, E=6.3e-20, E=1	317	27
1816	gi557822	Saccharomyces cerevisiae	mal5, sta1, len: 1367, CAI: 0.3, AMYH_YEAST P08640	311	29

223

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			GLUCOAMYLASE S1 (EC 3.2.1.3)		
1817	gi2432007	Mus musculus	peripheral benzodiazepine receptor associated protein; PBR associated protein; PAP20	178	67
1817	gi4490289	Human adenovirus type 41	33K protein	80	33
1817	gi12214169	Enterococcus hirae	penicillin binding protein 1A	95	28
1818	gi14042046	Homo sapiens	cDNA FLJ14492 fis, clone MAMMA1002937, weakly similar to ZINC FINGER PROTEIN 135.	3402	99
1818	AAB95103	Homo sapiens	Human protein sequence SEQ ID NO:17076.	3402	99
1818	gi10435504	Homo sapiens	cDNA FLJ13479 fis, clone PLACE1003738, weakly similar to ZINC FINGER PROTEIN 135.	2739	99
1819	AAG01908	Homo sapiens	Human secreted protein, SEQ ID NO: 5989.	307	100
1819	gi15025943	Clostridium acetobutylicum	D-alanine-D-alanine ligase	49	24
1819	gi14247014	Staphylococcus aureus subsp. aureus Mu50	RNase HII	52	25
1820	gi13278819	Homo sapiens	clone MGC:2776 IMAGE:2959536, mRNA, complete cds.	639	100
1820	gi13278906	Homo sapiens	clone MGC:4440 IMAGE:2959536, mRNA, complete cds.	639	100
1820	AAG01158	Homo sapiens	Human secreted protein, SEQ ID NO: 5239.	502	100
1821	AAB67454	Homo sapiens	Amino acid sequence of a human chaperone polypeptide.	1097	99
1821	gi6179666	Schizosaccharom yces pombe	dnaJ protein	353	39
1821	gi9294116	Arabidopsis thaliana	dnaJ protein-like	346	40
1822	gi3860093	Homo sapiens	MDC-3.13 isoform 2 mRNA, complete cds.	655	60
1822	gi13529164	Homo sapiens	TNF-induced protein, clone MGC:12451 IMAGE:3997650, mRNA, complete cds.	655	60
1822	gi13937826	Homo sapiens	TNF-induced protein, clone MGC:12346 IMAGE:3930240, mRNA, complete cds.	655	60
1823	gi165780	Oryctolagus cuniculus	ubiquitin conjugating-protein	379	100
1823	gi207555	Rattus norvegicus	ubiquitin conjugating-protein	379	100
1823	gi1237240	Mus musculus	ubiquitin-conjugating enzym	379	100
1824	gi1769491	Homo sapiens	Human kruppel-related zinc finger protein (ZNF184) mRNA, partial cds.	691	33
1824	gi868160	Rattus norvegicus	Cys2/His2 zinc finger protein	567	37
1824	gi13752754	Homo sapiens	zinc finger 1111 mRNA, complete	545	37

224
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			cds.		
1825	gi2749771	Mus musculus	tsec-1	1122	72
1825	AAB63598	Homo sapiens	Human gastric cancer associated antigen protein sequence SEQ ID NO:960.	135	22
1825	gi167835	Dictyostelium discoideum	myosin heavy chain	155	24
1826	gi10437151	Homo sapiens	cDNA: FLJ21120 fis, clone CAS05691.	2545	99
1826	gi531204	Hydra vulgaris	myosin heavy chain	107	21
1826	gi530361	Trypanosoma brucei	I2 protein	103	18
1827	AAB36623	Homo sapiens	Human FLEXHT-45 protein sequence SEQ ID NO:45.	916	100
1827	gi4929683	Homo sapiens	CGI-107 protein mRNA, complete cds.	183	25
1827	gi7023866	Homo sapiens	cDNA FLJ11295 fis, clone PLACE1009721, weakly similar to MSF1 PROTEIN.	183	25
1828	gi13543189	Mus musculus	Similar to RIKEN cDNA 1300007L22 gene	760	62
1828	gi64552	Xenopus laevis	apical protein	305	33
1828	gi1773381	Homo sapiens	chromosome X clone U177G4, U152H5, U168D5, 174A6, U172D6, and U186B3 from Xp22, complete sequence.	305	38
1829	AAG01378	Homo sapiens	Human secreted protein, SEQ ID NO: 5459.	575	100
1829	gi14250426	Homo sapiens	clone IMAGE:3866238, mRNA, partial cds.	109	25
1829	gi10437591	Homo sapiens	cDNA: FLJ21480 fis, clone COL05034.	99	29
1830	gi13752754	Homo sapiens	zinc finger 1111 mRNA, complete cds.	867	53
1830	gi10436789	Homo sapiens	cDNA FLJ14345 fis, clone THYRO1001189, weakly similar to ZINC FINGER PROTEIN 91.	862	56
1830	AAB95862	Homo sapiens	Human protein sequence SEQ ID NO:18929.	862	56
1832	AAB01386	Homo sapiens	Neuron-associated protein.	1561	98
1832	AAW29648	Homo sapiens	Human secreted protein BI164 1.	905	100
1832	gi9622219	Rattus norvegicus	beta-catenin binding protein	259	34
1833	gi432544	Trichoplusia ni, larvae, Peptide, 256 aa	HSUP59=28 kda trypsin protease homolog	74	41
1833	gi2618994	Bacillus subtilis	YoxB	71	37
1833	gi2634246	Bacillus subtilis	yoxB	71	37
1834	gi2394174	Homo sapiens	zinc finger protein ZNF191 (ZNF191) gene, complete cds.	818	56
1834	gi4405797	Homo sapiens	retinoic acid suppression protein A (RSG-A) mRNA, complete cds.	818	56
1834	gi13097726	Homo sapiens	zinc finger protein 24 (KOX 17), clone MGC:2057 IMAGE:3537160, mRNA, complete cds.	818	56

225
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1835	gi13623571	Homo sapiens	clone IMAGE:4309033, mRNA, partial cds.	1639	99
1835	gi11967437	Caenorhabditis elegans	suppressor of presenilin 2	81	35
1835	gi3425860	Streptomyces coelicolor	ScbA protein	102	28
1836	gi13623571	Homo sapiens	clone IMAGE:4309033, mRNA, partial cds.	705	99
1836	gi7649887	Escherichia coli O157:H7	tail fiber protein	99	37
1836	gi11967437	Caenorhabditis elegans	suppressor of presenilin 2	81	35
1837	gi10437973	Homo sapiens	cDNA: FLJ21801 fis, clone HEP00707.	468	77
1837	gi12052965	Homo sapiens	mRNA; cDNA DKFZp566M1046 (from clone DKFZp566M1046); complete cds.	188	26
1837	gi10439143	Homo sapiens	cDNA: FLJ22665 fis, clone HSI08219.	188	26
1838	gi13940223	Homo sapiens	partial mRNA for doublesex-mab-3 (DM) domain (DMRTA2 gene).	504	46
1838	gi9743439	Oreochromis niloticus	sex-determining protein DMO	408	43
1838	gi6466208	Mus musculus	doublesex and mab-3 related transcription factor 1	330	59
1839	gi10433552	Homo sapiens	cDNA FLJ12145 fis, clone MAMMA1000395.	2946	99
1839	AAB93901	Homo sapiens	Human protein sequence SEQ ID NO:13855.	2946	99
1839	gi12804489	Homo sapiens	clone MGC:2771 IMAGE:2958853, mRNA, complete cds.	1797	77
1840	gi12804489	Homo sapiens	clone MGC:2771 IMAGE:2958853, mRNA, complete cds.	3096	99
1840	gi13647045	Mus musculus	Slp homologue lacking C2 domains-a	1845	63
1840	gi10433552	Homo sapiens	cDNA FLJ12145 fis, clone MAMMA1000395.	1791	77
1841	gi12804489	Homo sapiens	clone MGC:2771 IMAGE:2958853, mRNA, complete cds.	1167	98
1841	gi10433552	Homo sapiens	cDNA FLJ12145 fis, clone MAMMA1000395.	1771	94
1841	AAB93901	Homo sapiens	Human protein sequence SEQ ID NO:13855.	1771	94
1842	gi12804489	Homo sapiens	clone MGC:2771 IMAGE:2958853, mRNA, complete cds.	1788	100
1842	gi10433552	Homo sapiens	cDNA FLJ12145 fis, clone MAMMA1000395.	2090	99
1842	AAB93901	Homo sapiens	Human protein sequence SEQ ID NO:13855.	2090	99
1843	gi10434777	Homo sapiens	cDNA FLJ12983 fis, clone NT2RP3000002.	556	100
1843	AAB94395	Homo sapiens	Human protein sequence SEQ ID NO:14962.	556	100
1843	gi2621988	Methanothermobacter	integrase-recombinase protein	77	31

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		thermautotrophic us			
1844	gi10435272	Homo sapiens	cDNA FLJ13305 fis, clone OVARC1001399.	386	35
1844	AAB94564	Homo sapiens	Human protein sequence SEQ ID NO:15341.	386	35
1844	gi9916	Plasmodium falciparum	liver stage antigen	162	21
1845	gi11994722	Arabidopsis thaliana	contains similarity to calmodulin~gene_id:F5N5.10	151	24
1845	gi14625425	Nicotiana tabacum	calmodulin NtCaM13	141	27
1845	gi535428	Pisum sativum	calmodulin-like protein	137	24
1846	gi7022811	Homo sapiens	cDNA FLJ10649 fis, clone NT2RP2005835, weakly similar to SHP1 PROTEIN.	403	39
1846	AAB93031	Homo sapiens	Human protein sequence SEQ ID NO:11803.	403	39
1846	gi2285790	Rattus norvegicus	p47	386	44
1847	gi10435522	Homo sapiens	cDNA FLJ13490 fis, clone PLACE1004118.	2317	99
1847	AAB94676	Homo sapiens	Human protein sequence SEQ ID NO:15624.	2317	99
1847	AAR70784	Homo sapiens	Saposin-C.	77	30
1848	AAAY59712	Homo sapiens	Secreted protein 33-54-1-B9-FL1.	398	100
1848	gi8979275	Chlamydophila pneumoniae J138	muramoylalanine-glutamate ligase	84	26
1848	gi7189877	Chlamydophila pneumoniae AR39	UDP-N-acetylmuramoylalanine--D-glutamate ligase	84	26
1849	gi9800509	Homo sapiens	pinch-2 protein mRNA, complete cds.	631	100
1849	AAG02299	Homo sapiens	Human secreted protein, SEQ ID NO: 6380.	428	98
1849	gi13542844	Mus musculus	RIKEN cDNA 4921524A02 gene	288	98
1850	gi10435023	Homo sapiens	cDNA FLJ13140 fis, clone NT2RP3003204.	2219	85
1850	AAB94491	Homo sapiens	Human protein sequence SEQ ID NO:15180.	2219	85
1850	AAAY73377	Homo sapiens	HTRM clone 1645941 protein sequence.	2213	100
1851	gi14250241	Mus musculus	Similar to prostate tumor over expressed gene 1	2151	96
1851	gi7920398	Homo sapiens	PTOV1 (PTOV1) gene, complete cds.	2150	98
1851	AAB58992	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 700.	1861	97
1852	gi14588846	Homo sapiens	titin zinc-finger anchoring protein, 50kDa isoform.	1954	96
1852	gi14588848	Homo sapiens	titin zinc-finger anchoring protein, 60kDa isoform.	1954	96
1852	gi14043532	Homo sapiens	clone MGC:12836	1876	96

227
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			IMAGE:4110783, mRNA, complete cds.		
1853	gi7022832	Homo sapiens	cDNA FLJ10661 fis, clone NT2RP2006106.	451	69
1853	AAB93042	Homo sapiens	Human protein sequence SEQ ID NO:11827.	451	69
1853	gi14603247	Homo sapiens	Similar to RIKEN cDNA 5730409G15 gene, clone MGC:19636 IMAGE:2822323, mRNA, complete cds.	254	74
1854	gi3387911	Homo sapiens	clone 24408 2-oxoglutarate carrier protein mRNA, complete cds.	1186	100
1854	gi13676350	Homo sapiens	clone MGC:2449 IMAGE:2960933, mRNA, complete cds.	1186	100
1854	gi13676368	Homo sapiens	clone MGC:4128 IMAGE:2960933, mRNA, complete cds.	1186	100
1855	gi10439606	Homo sapiens	cDNA: FLJ23042 fis, clone LNG02323.	676	96
1855	gi4699964	Homo sapiens	PAC clone RP5-953A4 from 7q11.23-q21.1, complete sequence.	107	24
1855	gi12240028	Myxococcus xanthus	translation initiation factor IF2	85	26
1856	gi6453487	Homo sapiens	mRNA; cDNA DKFZp434N0615 (from clone DKFZp434N0615); partial cds.	2276	100
1856	AAB56293	Homo sapiens	Human secreted protein sequence encoded by gene 99 SEQ ID NO:387.	715	67
1856	gi2230871	Homo sapiens	H.sapiens mRNA for Miz-1 protein.	502	30
1857	gi12232322	Homo sapiens	hUPF3A mRNA, complete cds.	2480	100
1857	gi12620406	Homo sapiens	UPF3 (UPF3) mRNA, complete cds.	2357	100
1857	AA Y32194	Homo sapiens	Human receptor molecule (REC) encoded by Incyte clone 266775.	1438	100
	gi12803415	Homo sapiens	clone MGC:2479 IMAGE:3140372, mRNA, complete cds.	611	100
	AAB58965	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 673.	500	96
	gi14330030	Mus musculus	bM401L17.7.1 (novel protein (isoform 1))	379	72
1858	AAT05577_aal	Homo sapiens	Human p69 cDNA.	544	50
1858	gi14250411	Homo sapiens	islet cell autoantigen 1 (69kD), clone MGC:9360 IMAGE:3854722, mRNA, complete cds.	544	50
1858	gi1674386	Homo sapiens	Human clone IS10 diabetes mellitus type I autoantigen (ICAp69) mRNA, complete cds.	544	50
1859	gi10434744	Homo sapiens	cDNA FLJ12960 fis, clone NT2RP2005605, weakly similar to QUEUINE TRNA-RIBOSYLTRANSFERASE (EC 2.4.2.29).	2182	100
1859	AAB94387	Homo sapiens	Human protein sequence SEQ ID NO:14945.	2182	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1859	gi6977808	Bacillus subtilis	tRNA-guanine transglycosylase	177	27
1860	gi10436039	Homo sapiens	cDNA FLJ13881 fis, clone THYRO1001458, moderately similar to MYOSIN HEAVY CHAIN, NONMUSCLE TYPE B.	2624	99
1860	AAB95605	Homo sapiens	Human protein sequence SEQ ID NO:18299.	2624	99
1860	gi12653781	Homo sapiens	clone IMAGE:3349352, mRNA, partial cds.	2266	96
1861	gi12053347	Homo sapiens	mRNA; cDNA DKFZp586M1120 (from clone DKFZp586M1120); complete cds.	1005	97
1861	gi4633066	Homo sapiens	protein phosphatase-1 regulatory subunit 7 (PPP1R7) gene, exon 11, complete cds and alternatively spliced product.	234	42
1861	gi4633068	Homo sapiens	protein phosphatase-1 regulatory subunit 7 (PPP1R7) gene, exons 8 and 9 and alternatively spliced products.	234	42
1862	gi7022518	Homo sapiens	cDNA FLJ10466 fis, clone NT2RP1001665.	488	38
1862	AAB92852	Homo sapiens	Human protein sequence SEQ ID NO:11417.	488	38
1862	gi10439049	Homo sapiens	cDNA: FLJ22601 fis, clone HSI04471.	858	99
1863	AAB70366	Homo sapiens	Human serine protease protein sequence SEQ ID NO:3.	1365	99
1863	gi6137097	Homo sapiens	serine protease DESC1 (DESC1) mRNA, complete cds.	660	48
1863	AAY99414	Homo sapiens	Human PRO1461 (UNQ742) amino acid sequence SEQ ID NO:269.	660	48
1864	AAG02826	Homo sapiens	Human secreted protein, SEQ ID NO: 6907.	248	90
1864	gi12718793	Yarrowia lipolytica	ND1 protein	54	32
1864	AAB56955	Homo sapiens	Human prostate cancer antigen protein sequence SEQ ID NO:1533.	59	29
1865	gi4512103	Bos taurus	rab11 binding protein	1177	84
1865	gi6049150	Rattus norvegicus	WD-containing protein	1027	78
1865	gi7715587	Streptococcus pneumoniae	PspA	136	26
1866	gi10439164	Homo sapiens	cDNA: FLJ22679 fis, clone HSI10687.	1365	100
1866	gi59578	Human herpesvirus 4	6.5kd unidentified reading frame (aa 1-60)	62	34
1866	gi1334920	Human herpesvirus 4	BNLF2a reading frame	62	34
1867	gi12652919	Homo sapiens	clone MGC:2803 IMAGE:2961319, mRNA, complete cds.	927	100
1867	AAY59763	Homo sapiens	Human normal ovarian tissue derived protein 40.	877	100
1867	AAB92974	Homo sapiens	Human protein sequence SEQ ID NO:11683.	551	100

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1868	AAG03616	Homo sapiens	Human secreted protein, SEQ ID NO: 7697.	429	96
1868	gi11093536	Homo sapiens	interleukin 21 (IL21) mRNA, complete cds.	75	25
1868	AAB18623	Homo sapiens	A human zalp11 ligand polypeptide.	75	25
1869	gi2114473	Mus musculus	p140mDia	488	29
1869	AAW76734	Homo sapiens	Human mDia Rho targeting protein.	490	28
1869	gi2947238	Homo sapiens	diaphanous 1 (HDIA1) mRNA, complete cds.	484	29
1870	gi12804143	Homo sapiens	clone IMAGE:3953631, mRNA, partial cds.	2777	99
1870	gi4432833	Arabidopsis thaliana	En/Spm-like transposon protein	110	29
1870	gi12804045	Homo sapiens	clone IMAGE:3940843, mRNA, partial cds.	116	28
1871	gi13603845	Mus musculus	ribonuclease/angiogenin inhibitor 2	1516	49
1871	gi11096305	Homo sapiens	NALP2 mRNA, complete cds.	802	33
1871	gi7020664	Homo sapiens	cDNA FLJ20510 fis, clone KAT09662.	802	33
1872	gi7023341	Homo sapiens	cDNA FLJ10968 fis, clone PLACE1000863, moderately similar to PUTATIVE MITOCHONDRIAL 40S RIBOSOMAL PROTEIN YHR148W.	536	61
1872	gi15079901	Homo sapiens	Similar to mitochondrial ribosomal protein S4, clone MGC:19665 IMAGE:3344519, mRNA, complete cds.	536	61
1872	gi12053321	Homo sapiens	mRNA; cDNA DKFZp586L0118 (from clone DKFZp586L0118); complete cds.	536	61
1873	AAV73467	Homo sapiens	Human secreted protein clone yd61_1 protein sequence SEQ ID NO:156.	494	98
1873	AAB45476	Homo sapiens	Human secreted protein BLAST search protein SEQ ID NO: 150.	366	97
1873	gi13905260	Mus musculus	RIKEN cDNA 1300006C06 gene	246	50
1874	gi14044001	Homo sapiens	clone MGC:14173 IMAGE:4120858, mRNA, complete cds.	2913	99
1874	gi10121792	Homo sapiens	MJ0495-like protein SelB mRNA, partial cds.	2704	99
1874	gi10121790	Mus musculus	MJ0495-like protein SelB	2662	87
1875	gi10434992	Homo sapiens	cDNA FLJ13119 fis, clone NT2RP3002671, weakly similar to ELONGATION FACTOR 2.	4440	99
1875	AAB94484	Homo sapiens	Human protein sequence SEQ ID NO:15164.	4440	99
1875	AAB07853	Homo sapiens	Amino acid sequence of Smad1 interactor protein clone S1+28.	2574	99
1876	gi12621066	Homo sapiens	RNA helicase-DEAD box protein RH116 mRNA, complete cds.	5309	100
1876	gi11344594	Homo sapiens	melanoma differentiation associated protein-5 (MDA5) mRNA, complete	5285	99

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			cds.		
1876	gi10432595	Homo sapiens	cDNA FLJ11354 fis, clone HEMBA1000129, weakly similar to HYPOTHETICAL HELICASE C8A4.08C IN CHROMOSOME I.	684	40
1877	gi4929721	Homo sapiens	CGI-126 protein mRNA, complete cds.	613	100
1877	gi6841532	Homo sapiens	HSPC155	613	100
1877	AAW74783	Homo sapiens	Human secreted protein encoded by gene 54 clone HMADJ02.	613	100
1878	gi2230873	Homo sapiens	H.sapiens mRNA for M phase phosphoprotein 10.	3398	99
1878	AAB34433	Homo sapiens	Human secreted protein sequence encoded by gene 46 SEQ ID NO:194.	2096	100
1878	AAB63275	Homo sapiens	Human breast cancer associated antigen protein sequence SEQ ID NO:637.	804	98
1879	gi12053035	Homo sapiens	mRNA; cDNA DKFZp434G1415 (from clone DKFZp434G1415); complete cds.	1134	100
1879	gi7020619	Homo sapiens	cDNA FLJ20485 fis, clone KAT07835.	241	31
1879	gi2622648	Methanothermobacter thermotrophicus	conserved protein	103	29
1880	gi506502	Mus musculus	NK10	276	68
1880	gi10440081	Homo sapiens	cDNA: FLJ23404 fis, clone HEP18862.	250	64
1880	gi7022523	Homo sapiens	cDNA FLJ10469 fis, clone NT2RP2000008, weakly similar to ZINC FINGER PROTEIN 84.	258	71
1881	gi10440371	Homo sapiens	mRNA for FLJ00020 protein, partial cds.	5999	99
1881	gi15077006	Homo sapiens	AT-hook protein AKNA mRNA, complete cds.	2454	79
1881	AAG73509	Homo sapiens	Human gene 35-encoded secreted protein fragment, SEQ ID NO:285.	944	100
1882	gi10433398	Homo sapiens	cDNA FLJ12015 fis, clone HEMBB1001695.	778	100
1882	AAB95018	Homo sapiens	Human protein sequence SEQ ID NO:16726.	778	100
1882	gi6562363	Canis familiaris	brain-specific synapse associated protein, Bassoon	71	40
1883	gi11527289	Homo sapiens	LHX5 protein mRNA, complete cds.	2174	100
1883	gi531220	Rattus norvegicus	amino acid feature: homeodomain, bp 895 .. 1074; amino acid feature: LIM1, bp 373 .. 516; amino acid feature: LIM2, bp 550 .. 705	2150	98
1883	gi1388183	Mus musculus	LIM/homeodomain	2150	98
1884	AAG01575	Homo sapiens	Human secreted protein, SEQ ID NO: 5656.	341	82
1884	gi14714594	Homo sapiens	clone MGC:15955 IMAGE:3538218, mRNA, complete	56	32

231
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			cds.		
1884	gi1340146	Homo sapiens	Human gene for L apoferritin exons 3 and 4.	53	34
1885	gi11602755	Mus musculus	zinc finger protein	1227	70
1885	gi11558482	Homo sapiens	mRNA for B-cell lymphoma/leukaemia 11A extra long form (BCL11A-XL gene).	300	40
1885	gi11558488	Homo sapiens	mRNA for B-cell lymphoma/leukaemia 11B (BCL11B gene).	298	46
1886	gi6705987	Mus musculus	phospholipase C-L2	270	39
1886	gi1183844	Rattus norvegicus	130kDa-Ins(1,4,5)P3 binding protein	266	40
1886	gi780122	Homo sapiens	Human mRNA for phospholipase C, complete cds.	264	40
1887	gi10436530	Homo sapiens	cDNA FLJ14146 fis, clone MAMMA1002947.	716	100
1887	AAB95718	Homo sapiens	Human protein sequence SEQ ID NO:18582.	716	100
1887	gi9928335	Mycobacterium tuberculosis	SEQ ID N_ 5A	68	33
1888	gi1542813	Mus musculus	Six5	2908	85
1888	gi6230605	Homo sapiens	SIX4 gene, exon 3 and complete cds.	806	52
1888	gi1255626	Mus musculus	AREC3	798	46
1889	gi10242353	Homo sapiens	pellino 2 (PELI2) mRNA, complete cds.	1636	70
1889	gi14550457	Homo sapiens	pellino (Drosophila) homolog 2, clone MGC:15066 IMAGE:3942712, mRNA, complete cds.	1636	70
1889	gi10242355	Mus musculus	pellino 1	1631	70
1890	gi12052983	Homo sapiens	mRNA; cDNA DKFZp434I1610 (from clone DKFZp434I1610); complete cds.	4179	100
1890	gi6467206	Homo sapiens	GIOT-4 mRNA for gonadotropin inducible transcription repressor-4, complete cds.	2071	57
1890	AAB94388	Homo sapiens	Human protein sequence SEQ ID NO:14947.	1976	54
1891	gi12052983	Homo sapiens	mRNA; cDNA DKFZp434I1610 (from clone DKFZp434I1610); complete cds.	825	79
1891	gi5262560	Homo sapiens	mRNA; cDNA DKFZp572P0920 (from clone DKFZp572P0920); partial cds.	460	56
1891	gi10434856	Homo sapiens	cDNA FLJ13032 fis, clone NT2RP3001120, moderately similar to ZINC FINGER PROTEIN 136.	453	56
1892	gi10439174	Homo sapiens	cDNA: FLJ22686 fis, clone HSI10987.	2500	99
1892	gi6224683	Homo sapiens	unconventional myosin-15 mRNA, complete cds.	759	34
1892	gi6224685	Mus musculus	unconventional myosin-15	741	33
1893	gi6453502	Homo sapiens	mRNA; cDNA DKFZp434B1917	1639	99

232
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			(from clone DKFZp434B1917); partial cds.		
1893	gi14488277	Homo sapiens	aat-1 mRNA for AAT-1 alpha, complete cds.	486	100
1893	gi14488279	Homo sapiens	aat-1 mRNA for AAT-1 beta, complete cds.	152	100
1894	gi6453502	Homo sapiens	mRNA; cDNA DKFZp434B1917 (from clone DKFZp434B1917); partial cds.	1639	99
1894	gi14488277	Homo sapiens	aat-1 mRNA for AAT-1 alpha, complete cds.	486	100
1894	gi14488279	Homo sapiens	aat-1 mRNA for AAT-1 beta, complete cds.	152	100
1895	gi9963996	Danio rerio	eukaryotic translation initiation factor eIF4E-1	791	75
1895	gi306487	Homo sapiens	cap-binding protein mRNA, complete cds.	778	66
1895	AA Y78505	Homo sapiens	Human general translation initiation factor eIF4E amino acid sequence.	778	66
1896	gi10433134	Homo sapiens	cDNA FLJ11795 fis, clone HEMBA1006155.	1549	100
1896	AAB93831	Homo sapiens	Human protein sequence SEQ ID NO:13650.	1549	100
1896	AAB53278	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:818.	70	40
1897	gi10433134	Homo sapiens	cDNA FLJ11795 fis, clone HEMBA1006155.	1558	100
1897	AAB93831	Homo sapiens	Human protein sequence SEQ ID NO:13650.	1558	100
1897	AAB53278	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:818.	70	40
1898	gi9963806	Homo sapiens	zinc finger protein ZNF287 (ZNF287) mRNA, complete cds.	794	59
1898	gi11527849	Mus musculus	zinc finger protein SKAT2	776	60
1898	gi8099348	Homo sapiens	zinc finger protein (ZFP) mRNA, complete cds.	787	44
1899	gi1339918	Homo sapiens	mRNA for TR3beta, complete cds.	1778	100
1899	gi1813882	Homo sapiens	Human NAK1 mRNA for DNA binding protein, complete cds.	1585	100
1899	gi292834	Homo sapiens	Human TR3 orphan receptor mRNA, complete cds.	1577	99
1900	gi2276396	Homo sapiens	Human C2f mRNA, complete cds.	1235	100
1900	gi2289907	Mus musculus	C2F	1136	88
1900	gi12805095	Mus musculus	gene rich cluster, C2f gene	1136	88
1901	gi2276396	Homo sapiens	Human C2f mRNA, complete cds.	797	87
1901	gi2289907	Mus musculus	C2F	722	78
1901	gi12805095	Mus musculus	gene rich cluster, C2f gene	722	78
1902	gi2696611	Rattus norvegicus	RNA splicing-related protein	981	84
1902	gi9837385	Takifugu rubripes	retinitis pigmentosa GTPase regulator-like protein	181	25
1902	AAB43954	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1399.	150	25
1903	gi4191594	Homo sapiens	protein serine/threonine phosphatase 4 regulatory subunit 1 (PP4R1)	503	65

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			mRNA, complete cds.		
1903	AAB73355	Homo sapiens	Human mesangial cell meg-1 protein.	503	65
1903	gi5824546	Caenorhabditis elegans	M04C9.6c	238	43
1904	AAW59874	Homo sapiens	Amino acid sequence of the cDNA clone CAT-1 (HTXET53).	686	100
1904	AAV08326	Homo sapiens	Human granulysin P522 active fragment.	680	99
1904	gi35065	Homo sapiens	Human NKG5 mRNA, expressed in natural killer cells and T-cells.	676	100
1905	gi35065	Homo sapiens	Human NKG5 mRNA, expressed in natural killer cells and T-cells.	354	100
1905	AAR23732	Homo sapiens	Gene 519 cDNA derived peptide.	354	100
1905	AAW59874	Homo sapiens	Amino acid sequence of the cDNA clone CAT-1 (HTXET53).	354	100
1906	AAR23732	Homo sapiens	Gene 519 cDNA derived peptide.	327	69
1906	gi35065	Homo sapiens	Human NKG5 mRNA, expressed in natural killer cells and T-cells.	325	67
1906	AAW59874	Homo sapiens	Amino acid sequence of the cDNA clone CAT-1 (HTXET53).	325	67
1907	gi10439180	Homo sapiens	cDNA: FLJ22690 fis, clone HSI11134.	1109	100
1907	gi5817053	Homo sapiens	mRNA; cDNA DKFZp586D0824 (from clone DKFZp586D0824); partial cds.	356	45
1907	gi13569476	Mus musculus	immunity-associated nucleotide 4	354	44
1908	gi10434137	Homo sapiens	cDNA FLJ12571 fis, clone NT2RM4000950.	2885	99
1908	AAB94136	Homo sapiens	Human protein sequence SEQ ID NO:14401.	2885	99
1908	gi6807927	Homo sapiens	mRNA; cDNA DKFZp434I042 (from clone DKFZp434I042); partial cds.	1730	99
1909	gi10440532	Homo sapiens	mRNA for FLJ00116 protein, partial cds.	891	100
1909	gi14211720	Homo sapiens	desmuslin mRNA, complete cds.	104	34
1909	gi14588604	Oryza sativa	transcription factor rough sheath 2 like protein	79	32
1910	gi10440383	Homo sapiens	mRNA for FLJ00027 protein, partial cds.	1272	100
1910	gi14336762	Homo sapiens	16p13.3 sequence section 7 of 8.	1255	99
1910	gi10801127	Mus musculus	JNK/SAPK-associated protein 1d	1243	97
1911	gi4426613	Mus musculus	SLM-1	1780	95
1911	gi3822555	Mus musculus	SLM-2	1186	67
1911	gi3417603	Mus musculus	ETOILE	1178	67
1912	AAV28682	Homo sapiens	Human pp392_3 secreted protein.	2457	96
1912	AAG03379	Homo sapiens	Human secreted protein, SEQ ID NO: 7460.	377	100
1913	gi10434919	Homo sapiens	cDNA FLJ13074 fis, clone NT2RP3001855, moderately similar to HOMEBOX PROTEIN PKNOX1.	1543	98
1913	AAB94455	Homo sapiens	Human protein sequence SEQ ID NO:15102.	1543	98

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1913	gi14043520	Homo sapiens	clone MGC:12791 IMAGE:4099156, mRNA, complete cds.	1261	61
1914	gi10434841	Homo sapiens	cDNA FLJ13022 fis, clone NT2RP3000753, weakly similar to NEUROFILAMENT TRIPLET H PROTEIN.	4740	100
1914	AAB94420	Homo sapiens	Human protein sequence SEQ ID NO:15022.	4740	100
1914	gi4884180	Homo sapiens	mRNA; cDNA DKFZp564P1916 (from clone DKFZp564P1916); partial cds.	1129	100
1915	gi8925970	Homo sapiens	eIF4E-transporter mRNA, complete cds.	5059	99
1915	gi9857633	Mus musculus	Clast4 protein	4581	89
1915	gi10437728	Homo sapiens	cDNA: FLJ21601 fis, clone COL07221.	2997	100
1916	gi9857633	Mus musculus	Clast4 protein	4264	85
1916	gi8925970	Homo sapiens	eIF4E-transporter mRNA, complete cds.	4095	90
1916	gi10437728	Homo sapiens	cDNA: FLJ21601 fis, clone COL07221.	2033	83
1917	gi14714937	Homo sapiens	clone MGC:17687 IMAGE:3865868, mRNA, complete cds.	2052	100
1917	gi2443272	Mus musculus	motor domain of KIF12	611	88
1917	AAW27653	Homo sapiens	Secreted protein AS32.	516	83
1918	gi307166	Homo sapiens	Human mineralocorticoid receptor mRNA (hMR), complete cds.	523	79
1918	AAY21622	Homo sapiens	Ligand binding domain of nuclear receptor hMR.	523	79
1918	AAP80927	Homo sapiens	Sequence of the human mineralocorticoid receptor (hMR).	523	79
1919	gi1395179	Rattus norvegicus	proteasomal ATPase (MSS1)	922	100
1919	gi12803525	Homo sapiens	proteasome (prosome, macropain) 26S subunit, ATPase, 2, clone MGC:3004 IMAGE:3161790, mRNA, complete cds.	922	100
1919	gi219931	Homo sapiens	Human mRNA for MSS1, complete cds.	922	100
1921	gi300482	human, endogenous retroviral element RTVL-Hp1, Genomic, 660 nt]. [Homo sapiens	pol=reverse transcriptase homolog {retroviral element}	501	80
1921	gi199851	Mus musculus	pol protein	355	45
1921	gi927201	Murine leukemia virus	reverse transcriptase	355	45
1923	gi10862828	Homo sapiens	mRNA for IFRG15 protein.	723	99
1923	gi10862856	Mus musculus	15kD interferon alpha responsive protein	716	97
1923	gi7658271	Ginglymostoma	MHC class II protein beta chain	41	50

235
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		cirratum			
1924	gi495273	Rattus norvegicus	ribosomal protein S15a	565	84
1924	AAB53695	Homo sapiens	Human colon cancer antigen protein sequence SEQ ID NO:1235.	565	84
1924	AAB43652	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1097.	565	84
1925	gi15029945	Homo sapiens	clone MGC:17507 IMAGE:3455576, mRNA, complete cds.	1700	99
1925	gi7020140	Homo sapiens	cDNA FLJ20200 fis, clone COLF1206.	545	41
1925	gi12654477	Homo sapiens	clone IMAGE:2823731, mRNA, partial cds.	472	41
1926	gi55483	Mus musculus	Zfp-1 protein (AA 1-424)	1738	85
1926	gi15081398	Homo sapiens	kruppel-like zinc finger protein (ZNF300) mRNA, complete cds.	965	48
1926	gi1020145	Homo sapiens	Human DNA binding protein (HPF2) mRNA, complete cds.	925	53
1927	gi425551	Oryctolagus cuniculus	K3 keratin	1206	54
1927	gi433958	Oryctolagus cuniculus	keratin K3	1206	54
1927	gi14250682	Homo sapiens	keratin 6A, clone MGC:10443 IMAGE:3947610, mRNA, complete cds.	1267	53
1928	gi10437560	Homo sapiens	cDNA: FLJ21457 fis, clone COL04705.	2563	99
1928	AAB21000	Homo sapiens	Human nucleic acid-binding protein, NuABP-4.	1409	100
1928	gi7239109	Homo sapiens	HSPC059	1305	49
1929	gi10436789	Homo sapiens	cDNA FLJ14345 fis, clone THYRO1001189, weakly similar to ZINC FINGER PROTEIN 91.	3435	99
1929	AAB95862	Homo sapiens	Human protein sequence SEQ ID NO:18929.	3435	99
1929	gi10440398	Homo sapiens	mRNA for FLJ00032 protein, partial cds.	2509	80
1930	AAB43912	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1357.	1604	52
1930	gi1020145	Homo sapiens	Human DNA binding protein (HPF2) mRNA, complete cds.	1427	42
1930	gi12584159	Homo sapiens	zinc finger protein 268 (ZNF268) mRNA, complete cds.	1361	41
1931	gi13559680	Caenorhabditis elegans	contains similarity to beta-lactamases	144	25
1931	AAB58722	Homo sapiens	Breast and ovarian cancer associated antigen protein sequence SEQ ID 430.	80	25
1931	gi623764	Entamoeba histolytica	elongation factor 1a	72	28
1932	gi7022392	Homo sapiens	cDNA FLJ10390 fis, clone NT2RM4000104, moderately similar to ZINC FINGER PROTEIN 135.	440	43

236

Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1932	AAB92770	Homo sapiens	Human protein sequence SEQ ID NO:11245.	440	43
1932	gi15100039	Mus musculus	zinc finger protein ZFEND	438	43
1933	gi9963806	Homo sapiens	zinc finger protein ZNF287 (ZNF287) mRNA, complete cds.	4142	100
1933	gi11527849	Mus musculus	zinc finger protein SKAT2	3432	82
1933	gi2293535	Homo sapiens	zinc finger protein (ZnF20) mRNA, complete cds.	1366	43
1934	gi10439752	Homo sapiens	cDNA: FLJ23157 fis, clone LNG09620.	1525	99
1934	gi13603873	Homo sapiens	TBP-associated factor II Q (TAF2Q) mRNA, complete cds.	1177	96
1934	gi13603831	Mus musculus	TBP-associated factor II Q	670	63
1935	gi1060973	Oryctolagus cuniculus	Potassium channel	105	30
1935	gi5932024	Oryctolagus cuniculus	voltage-gated potassium channel Kv1.5	105	30
1935	gi3044206	Oryctolagus cuniculus	potassium channel Kv1.5	105	30
1936	gi4150892	Rattus norvegicus	dermo-1 protein	810	100
1936	gi1098934	Mus musculus	Dermo-1	810	100
1936	gi4200314	Gallus gallus	Dermo protein	782	96
1937	gi13752754	Homo sapiens	zinc finger 1111 mRNA, complete cds.	2304	58
1937	gi10440398	Homo sapiens	mRNA for FLJ00032 protein, partial cds.	2294	59
1937	gi10436789	Homo sapiens	cDNA FLJ14345 fis, clone THYRO1001189, weakly similar to ZINC FINGER PROTEIN 91.	2081	59
1938	gi14787181	Homo sapiens	CUB and sushi multiple domains protein 1 short form mRNA, complete cds, alternatively spliced.	789	58
1938	gi14794726	Homo sapiens	CUB and sushi multiple domains 1 protein mRNA, complete cds.	789	58
1938	gi14787176	Mus musculus	CSMD1	784	54
1939	gi14787181	Homo sapiens	CUB and sushi multiple domains protein 1 short form mRNA, complete cds, alternatively spliced.	256	56
1939	gi14794726	Homo sapiens	CUB and sushi multiple domains 1 protein mRNA, complete cds.	256	56
1939	gi14787176	Mus musculus	CSMD1	255	56
1940	gi12053079	Homo sapiens	mRNA; cDNA DKFZp434F0318 (from clone DKFZp434F0318); complete cds.	1212	98
1940	gi13384259	Homo sapiens	apolipoprotein L6 mRNA, complete cds.	134	31
1940	AAV35897	Homo sapiens	Extended human secreted protein sequence, SEQ ID NO. 146.	130	32
1941	gi12804053	Homo sapiens	clone IMAGE:3940029, mRNA, partial cds.	5182	99
1941	AAU00416	Homo sapiens	Human cell regulatory protein p193.	2623	66
1941	AAB43850	Homo sapiens	Human cancer associated protein sequence SEQ ID NO:1295.	954	61
1942	gi12654005	Homo sapiens	Similar to ribosomal protein S9,	840	100

237
Table 2A

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			clone MGC:5482 IMAGE:3452221, mRNA, complete cds.		
1942	gi13938523	Homo sapiens	clone MGC:2458 IMAGE:2964451, mRNA, complete cds.	840	100
1942	gi13938567	Homo sapiens	clone MGC:4138 IMAGE:2964451, mRNA, complete cds.	840	100

238
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
972	gi1817578	Homo sapiens	alpha-globin 2	600	85
972	gi1817577	Homo sapiens	alpha-globin 1	600	85
972	gi28558	Homo sapiens	reading frame alpha-globin	600	85
973	AAW62041	Homo sapiens	KYOW Human arginase.	1559	94
973	AAR05306	Homo sapiens	TOYJ Human arginase.	1559	94
973	gi1197498	Homo sapiens	arginase	1559	94
974	AAW62041	Homo sapiens	KYOW Human arginase.	1552	94
974	AAR05306	Homo sapiens	TOYJ Human arginase.	1552	94
974	gi1197498	Homo sapiens	arginase	1552	94
975	AAW62041	Homo sapiens	KYOW Human arginase.	1538	94
975	AAR05306	Homo sapiens	TOYJ Human arginase.	1538	94
975	gi1197498	Homo sapiens	arginase	1538	94
976	AAB93175	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12114.	4912	100
976	gi14211500	Homo sapiens	secretory protein SEC8	4912	100
976	gi1019441	Rattus norvegicus	rsec8	4690	95
977	AAB39308	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 48 SEQ ID NO:188.	954	76
977	gi14091475	Homo sapiens	hairy cell leukemia protein 1	954	76
977	gi11999092	Homo sapiens	TGF beta inducible nuclear protein TINP1	954	76
978	gi20198487	Homo sapiens	182kDa tankyrase1-binding protein	6194	99
978	gi18676574	Homo sapiens	FLJ00184 protein	2747	99
978	gi17864717	Mus musculus	hornerin	188	21
979	AAAY13459	Homo sapiens	UYRQ Amino acid sequence of human Fe65-like protein.	3643	96
979	gi1657752	Homo sapiens	FE65-like protein	3643	96
979	gi13377732	Rattus norvegicus	FE65	1556	46
980	AAAY13459	Homo sapiens	UYRQ Amino acid sequence of human Fe65-like protein.	3492	94
980	gi1657752	Homo sapiens	FE65-like protein	3492	94
980	gi13377732	Rattus norvegicus	FE65	1524	46
981	AAT92305_aa1	Homo sapiens	SALK Constitutively active receptor-alpha encoding cDNA.	437	98
981	AAG63170	Homo sapiens	TULA- Amino acid sequence of human CAR-a polypeptide.	437	98
981	AAW93902	Homo sapiens	GEHO Human CAR receptor protein.	437	98
982	gi19911227	Homo sapiens	2-amino-3-carboxylmuconate-6-semialdehyde decarboxylase	1438	95
982	gi18652911	Rattus norvegicus	2-amino-3-carboxymuconate-6-semialdehyde decarboxylase	1277	82
982	gi19911231	Caenorhabditis elegans	2-amino-3-carboxylmuconate-6-semialdehyde decarboxylase	660	42
983	AAM48092	Homo sapiens	SUMU Human IRG27 protein SEQ ID NO 1.	786	88
983	AAM38928	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2073.	786	88
983	AAB24220	Homo sapiens	SUMU Human immortalisation related gene 27 protein sequence SEQ ID NO:1.	786	88

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
984	gi20467066	Homo sapiens	thiamine triphosphatase	500	100
984	gi20467068	Bos taurus	thiamine triphosphatase	445	88
984	gi20467070	Mus musculus	thiamine triphosphatase	402	63
985	AAB93694	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13262.	764	49
985	AAM41100	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6031.	764	49
985	AAM39314	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2459.	764	49
986	AAO02579	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 16471.	1289	96
986	AAO02491	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 16383.	337	55
986	ABB11103	Homo sapiens	HYSE- Human tankyrase-1 homologue, SEQ ID NO:1473.	302	78
987	gi14456631	Homo sapiens	dJ54B20.4 (novel KRAB box containing C2H2 type zinc finger protein)	3439	100
987	AAU15940	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 893.	2016	100
987	gi1020145	Homo sapiens	DNA binding protein	1735	49
988	AAM95265	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 3923.	264	96
988	gi5901659	Caenorhabditis elegans	XNP-1	142	20
988	gi3253105	Caenorhabditis elegans	C. elegans XNP-1 protein (corresponding sequence B0041.7)	142	20
989	AAU00670	Homo sapiens	MILL- Human TANGO 229 polypeptide.	1020	60
989	gi20330504	Homo sapiens	CLCP1	440	31
989	gi20988615	Homo sapiens	Similar to endothelial and smooth muscle cell-derived neuropilin-like protein	440	31
990	AAB56613	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1191.	1120	100
990	gi182642	Homo sapiens	rapamycin-binding protein	1109	96
990	gi182644	Homo sapiens	FK506-binding protein 25	1061	100
991	AAU03587	Homo sapiens	INCY- Human DNA modification protein, DNAMP-2.	894	77
991	gi12746410	Mus musculus	coenzyme A diphosphatase	507	48
991	gi10764850	Arabidopsis thaliana	F1K23.5	117	31
992	AAB56638	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1216.	2381	91
992	gi517065	Homo sapiens	chaperonin-like protein	2381	91
992	gi14348900	Homo sapiens	heat shock protein	2381	91
993	gi3298472	Mus musculus	zinc finger protein	828	91
993	gi14329524	Homo sapiens	dJ871E2.1 (novel protein (ortholog of mouse zinc finger protein Zan75))	545	96
993	AAG75641	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6405.	364	97
994	AAB95791	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18751.	3362	94
994	AAB95291	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17513.	3359	93

240
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
994	AAM93796	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3827.	3351	93
995	gi2736284	Mus musculus	Ldb1a	2224	99
995	gi2582522	Gallus gallus	neural src interacting protein, long form; NSIP long form	2206	98
995	gi5123791	Homo sapiens	Nuclear LIM interactor	2175	100
996	AAB62202	Homo sapiens	RIGE- Cell cycle protein Radh-isoform 2.	617	100
996	AAB62201	Homo sapiens	RIGE- Cell cycle protein Radh-isoform 1.	617	100
996	gi20072481	Mus musculus	Similar to RIKEN cDNA 1700019D06 gene	547	60
997	AAU75578	Homo sapiens	UYNA- Human ubiquitin specific protease 10 (USP10).	6817	100
997	gi13560797	Homo sapiens	ubiquitin specific protease	6817	100
997	AAM40478	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5409.	4801	99
998	AAE13285	Homo sapiens	INCY- Human transporters and ion channels (TRICH)-12.	511	62
998	gi 2072966 gb AAC51272.1	Homo sapiens	p40	366	42
999	gi17429686	Ralstonia solanacearum	PUTATIVE D-LACTATE DEHYDROGENASE (CYTOCHROME) OXIDOREDUCTASE PROTEIN	1217	51
999	gi17982522	Brucella melitensis	D-LACTATE DEHYDROGENASE (CYTOCHROME)	1148	47
999	gi17740510	Agrobacterium tumefaciens str. C58 (U. Washington)	FAD dependent oxidoreductase	1147	48
1000	AAB36613	Homo sapiens	INCY- Human FLEXHT-35 protein sequence SEQ ID NO:35.	528	71
1000	gi14603247	Homo sapiens	Similar to RIKEN cDNA 5730409G15 gene	528	71
1000	AAB93042	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11827.	220	84
1001	AAO10631	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 24523.	186	72
1001	gi13421401	Caulobacter crescentus CB15	MutT/nudix family protein	133	40
1001	ABB05692	Homo sapiens	GEHU- Human amygdala derived protein clone amy2_2i17.	112	29
1002	gi4689229	Rattus norvegicus	b-tomosyn isoform	2075	94
1002	gi3790389	Rattus norvegicus	m-tomosyn	1976	95
1002	gi4689231	Rattus norvegicus	s-tomosyn isoform	1896	94
1003	AAM40991	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5922.	365	83
1003	AAM39205	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2350.	356	82
1003	AAO07159	Homo sapiens	HYSE- Human polypeptide SEQ ID	328	75

241
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			NO 21051.		
1004	AAY21631	Homo sapiens	REGC Ligand binding domain of nuclear receptor hTRbeta.	524	100
1004	AAP80921	Homo sapiens	SALK Sequence encoded by human placenta thyroid receptor c-erbA cDNA.	497	100
1004	gi31207	Homo sapiens	put.thyroid hormone receptor	497	100
1005	gi3192954	Homo sapiens	ataxin-7	1065	35
1005	gi2370155	Homo sapiens	spinocerebellar ataxia 7	1064	35
1005	gi19071468	Mus musculus	ataxin-7	1020	35
1006	gi3192954	Homo sapiens	ataxin-7	1250	40
1006	gi2370155	Homo sapiens	spinocerebellar ataxia 7	1249	40
1006	gi19071468	Mus musculus	ataxin-7	1186	40
1007	AAU16630	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1583.	1106	97
1007	AAU16631	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1584.	970	99
1007	AAU16223	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1176.	947	99
1008	gi38482	Homo sapiens	proteasome subunit LMP7	1399	83
1008	gi1054747	Homo sapiens	alternative first exon (1b)	1399	83
1008	gi596140	Homo sapiens	proteasome subunit LMP7	1369	82
1009	AAB93756	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13429.	2396	96
1009	gi14165506	Homo sapiens	DNA methyltransferase 1-associated protein 1	2396	96
1009	gi9309471	Homo sapiens	DNMT1 associated protein-1	2396	96
1010	gi6942267	Danio rerio	bHLH transcription factor Mesp-b	78	27
1010	gi1730260	Bos taurus	aggrecan	76	26
1010	gi3395774	Aplysia californica	G-protein-coupled 5-hydroxytryptamine receptor	72	29
1011	AAU17585	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 1150.	1010	86
1011	AAU17212	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 777.	1010	86
1011	gi5815353	Homo sapiens	J domain containing protein 1 isoform a	1010	86
1012	gi16741189	Homo sapiens	Similar to nuclear prelamin A recognition factor	2143	100
1012	AAB92498	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10602.	2084	89
1012	AAB97260	Homo sapiens	SHAN- Human NADP hydrogenase subunit 50.	2084	89
1013	gi5917666	Zea mays	extensin-like protein	451	27
1013	gi15145793	Sus scrofa	basic proline-rich protein	422	27
1013	gi600118	Zea mays	extensin-like protein	394	26
1014	gi407468	Mus musculus	SEB4	1166	93
1014	gi407419	Homo sapiens	SEB4D	1141	96
1014	gi19851930	Homo sapiens	CLL-associated antigen KW-5	1134	100
1015	AAG81328	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:174.	1562	100
1015	AAB75532	Homo sapiens	ROSE/ Human secreted protein sequence encoded by gene 27 SEQ ID NO:86.	680	92

242
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1015	gi9802536	Arabidopsis thaliana	F17L21.25	305	24
1016	AAB43450	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:895.	779	98
1016	AAM69101	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29407.	132	100
1016	AAM56723	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 28828.	132	100
1017	AAB43450	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:895.	723	76
1017	AAM69101	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29407.	132	100
1017	AAM56723	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 28828.	132	100
1018	gi12653673	Homo sapiens	tubulin, gamma 1	2329	97
1018	gi183703	Homo sapiens	gamma-tubulin	2320	96
1018	gi4115724	Rattus norvegicus	tubulin	2304	96
1019	gi12653689	Homo sapiens	Similar to aspartyl-tRNA synthetase	2586	100
1019	gi179102	Homo sapiens	aspartyl-tRNA synthetase	2545	98
1019	gi14250408	Mus musculus	Similar to aspartyl-tRNA synthetase	2521	96
1020	gi18448004	Homo sapiens	nuclear associated protein	2867	100
1020	AAM93729	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3689.	2859	99
1020	gi18460918	Mus musculus	nuclear protein NAP	2756	95
1021	AAU91297	Homo sapiens	CORT- Human protein NOV4.	1744	100
1021	AAB73671	Homo sapiens	INCY- Human oxidoreductase protein ORP-4.	1744	100
1021	AAU79001	Homo sapiens	CORT- Human NOV4 variant protein.	1743	99
1022	ABB04369	Homo sapiens	MILL- Human 32225 polypeptide.	1834	100
1022	AAB94324	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14807.	1819	99
1022	AAG67147	Homo sapiens	INCY- Amino acid sequence of a human enzyme.	1031	55
1023	AAM58143	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 30248.	612	100
1023	gi17861968	Drosophila melanogaster	LD03052p	197	50
1023	gi4996365	Arabidopsis thaliana	polyprotein	78	31
1024	AAB93824	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13634.	719	100
1024	gi18605620	Mus musculus	similar to proline-rich protein 48	86	38
1024	gi5902891	Streptomyces avermitilis	type I polyketide synthase AVES 1	81	32
1025	ABB50184	Homo sapiens	INCY- Human transcription factor TRFX-35.	1452	57
1025	AAB93164	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12091.	1372	58
1025	AAM40308	Homo sapiens	HYSE- Human polypeptide SEQ ID	1358	49

243
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			NO 3453.		
1026	gi11386005	Homo sapiens	hepatocellular carcinoma-associated protein HCA10	2802	100
1026	gi17981589	Homo sapiens	EH domain-containing protein-4	2802	100
1026	gi13021980	Homo sapiens	hepatocellular carcinoma-associated protein HCA11	2802	100
1027	AAB64400	Homo sapiens	INCY- Amino acid sequence of human intracellular signalling molecule INTRA32.	1238	95
1027	AAB94943	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16446.	1238	95
1027	AAB93989	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14087.	1238	95
1028	AAB64400	Homo sapiens	INCY- Amino acid sequence of human intracellular signalling molecule INTRA32.	1120	88
1028	AAB94943	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16446.	1120	88
1028	AAB93989	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14087.	1120	88
1029	gi12697477	Homo sapiens	dJ850E9.2 (novel protein similar to Drosophila CG6762)	684	100
1029	gi16877494	Homo sapiens	Similar to RIKEN cDNA 1700127B04 gene	678	100
1029	gi15030144	Mus musculus	Similar to RIKEN cDNA 1700127B04 gene	613	88
1030	gi20987276	Mus musculus	RIKEN cDNA 5730409F23 gene	1778	99
1030	gi1381027	Homo sapiens	phosphoribosylpyrophosphate synthetase-associated protein 39	1775	99
1030	gi436779	Rattus norvegicus	phosphoribosylpyrophosphate synthetase-associated protein (39 kDa)	1768	98
1031	AAB35408	Homo sapiens	HOSP- Human 07CG27 gene protein.	619	28
1031	AAB93993	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14095.	411	28
1031	AAO04356	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 18248.	377	98
1032	AAB95299	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17530.	2238	99
1032	AAB67054	Homo sapiens	INCY- Human immune response molecule (IMUN) protein SEQ ID NO: 8.	1414	98
1032	AAM95465	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4123.	735	91
1033	gi15809588	Homo sapiens	hnRNP Q2	3140	99
1033	AAG74545	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5309.	3100	94
1033	AAV70242	Homo sapiens	INCY- Human RNA-associated protein-23 (RNAAP-23).	3100	94
1034	gi285933	Homo sapiens	Mel-18 protein	1811	97
1034	gi18999362	Homo sapiens	zinc finger protein 144 (Mel-18)	1811	97
1034	gi19548760	Mus musculus	mel-18 protein	1738	93
1035	AAZ21227_aa1	Homo sapiens	UYUP- Human CG1CE short form cDNA sequence.	859	53
1035	gi3598876	Homo sapiens	vitelliform macular dystrophy protein	859	53
1035	gi3335159	Homo sapiens	bestrophin	859	53

244
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1036	gi 19526874 [ref NP_598464.1]	Mus musculus	RIKEN cDNA 1500026F15	208	66
1036	gi 17550864 [ref NP_509196.1]	Caenorhabditis elegans	C36B7.6.p	70	33
1037	AAM41409	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6340.	994	99
1037	AA48501	Homo sapiens	META- Human breast tumour-associated protein 46.	987	98
1037	AAB67445	Homo sapiens	INCY- Amino acid sequence of a human chaperone polypeptide.	869	100
1038	AAM83952	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:11545.	572	97
1038	AAG01992	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6073.	296	100
1038	AAI65790_aa1	Homo sapiens	GBFB Nucleotide sequence of a human Fyb/SLAP1 protein.	76	28
1039	AAU20513	Homo sapiens	HUMA- Human secreted protein, Seq ID No 505.	1348	99
1039	gi20379968	Mus musculus	similar to FLJ00101 protein	140	33
1039	gi10440510	Homo sapiens	FLJ00101 protein	139	36
1040	AAU20513	Homo sapiens	HUMA- Human secreted protein, Seq ID No 505.	959	77
1040	gi20379968	Mus musculus	similar to FLJ00101 protein	140	33
1040	gi10440510	Homo sapiens	FLJ00101 protein	139	36
1041	gi12652649	Homo sapiens	ribosomal protein L28	638	76
1041	gi488835	Mus musculus	ribosomal protein L28	631	75
1041	gi57113	Rattus norvegicus	ribosomal protein L28 (AA 1-137)	611	73
1042	ABB11796	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2166.	4262	99
1042	AAE07171	Homo sapiens	PHAR- Human HuIFRG-3 protein.	4249	99
1042	AAM25341	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:856.	2267	99
1043	AAB43873	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1318.	1004	91
1043	gi306809	Homo sapiens	glutathione S-transferase	1004	91
1043	gi306815	Homo sapiens	glutathione S-transferase (GST, EC 2.5.1.18)	1004	91
1044	AAB43873	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1318.	990	90
1044	gi306809	Homo sapiens	glutathione S-transferase	990	90
1044	gi306815	Homo sapiens	glutathione S-transferase (GST, EC 2.5.1.18)	990	90
1045	AAB43873	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1318.	1008	91
1045	gi306809	Homo sapiens	glutathione S-transferase	1008	91
1045	gi306815	Homo sapiens	glutathione S-transferase (GST, EC 2.5.1.18)	1008	91
1046	AAM90773	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:18366.	905	97
1046	AAB93267	Homo sapiens	HELI- Human protein sequence SEQ	753	43

245
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			ID NO:12300.		
1046	gi 20863434 [ref]XP_125 026.1]	Mus musculus	similar to expressed sequence C85457	4581	85
1047	gi520450	Homo sapiens	sorbitol dehydrogenase	1733	94
1047	gi496086	Homo sapiens	L-idoitol-2 dehydrogenase	1726	94
1047	AAB56748	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1326.	1622	94
1048	AAB58331	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 669.	478	97
1048	gi6066468	Leishmania major	probable DNA polymerase zeta catalytic component	82	32
1048	AAM88708	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:16301.	77	43
1049	AAM49036	Homo sapiens	BODE- Human phosphatidylinositol 3 (PtdIns 3)-kinase 36.	1633	100
1049	AAB95783	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18735.	1626	99
1049	AAM94021	Homo sapiens	HELI- Human stomach cancer expressed polypeptide SEQ ID NO 112.	1626	99
1050	AAB56131	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 55 SEQ ID NO:225.	948	95
1050	gi6841554	Homo sapiens	HSPC166	940	95
1050	gi20987916	Mus musculus	RIKEN cDNA 1810029F08 gene	849	86
1051	AAB94709	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15705.	1819	99
1051	gi18447408	Drosophila melanogaster	RE13191p	144	29
1051	gi15809236	Plasmodium falciparum	merozoite surface protein 2	117	38
1052	gi15559064	Mus musculus	SNAG1	373	56
1052	gi6164628	Homo sapiens	SH3 and PX domain-containing protein SH3PX1	259	42
1052	gi5410249	Homo sapiens	SDP1 protein	259	42
1053	gi21410770	Homo sapiens	Similar to RIKEN cDNA 1500005K14 gene	1137	100
1053	gi18958091	Danio rerio	SI:bZ71M17.1 (novel protein similar to mouse proteins)	384	40
1053	AAM93224	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2637.	149	43
1054	AAM79121	Homo sapiens	HYSE- Human protein SEQ ID NO 1783.	2111	95
1054	gi28384	Homo sapiens	5' half of the product is homologues to Bacillus subtilis SAICAR synthetase, 3' half corresponds to the catalytic subunit of AIR carboxylase	2111	95
1054	gi17939425	Homo sapiens	multifunctional polypeptide similar to SAICAR synthetase and AIR carboxylase	2111	95
1055	AAG01499	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5580.	293	100
1055	gi 20875035	Mus musculus	similar to RNA polymerase II	74	25

246
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	ref XP_131118.1		termination factor		
1056	gi17945783	Drosophila melanogaster	RE34039p	278	36
1056	gi2072290	Xenopus laevis	XL-INCENP	113	25
1056	gi21038841	Caenorhabditis elegans	Y39B6A.18	113	25
1057	AAO13523	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 27415.	1302	83
1057	AAB95348	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17635.	171	29
1057	gi14550467	Homo sapiens	spermatogenesis associated 2	171	29
1058	AAM52672	Homo sapiens	SHAN- Human microglobulin transcriptional control factor 30.	1172	87
1058	gi16588712	Homo sapiens	P33	1172	87
1058	gi14250169	Homo sapiens	Similar to leucine zipper protein FKSG14	1172	87
1059	AAO16945	Homo sapiens	KYOW Human NF-kappaB activity enhancing protein SEQ ID NO: 1.	2622	100
1059	AAO16949	Homo sapiens	KYOW Human NF-kappaB activity enhancing protein SEQ ID NO: 5.	2608	99
1059	AAB62180	Homo sapiens	PLAC Human p95 protein.	2395	89
1060	gi16648230	Drosophila melanogaster	GH23825p	329	33
1060	gi13186114	Homo sapiens	rab interacting lysosomal protein	200	30
1060	AAG02093	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6174.	136	41
1061	gi10440398	Homo sapiens	FLJ00032 protein	2225	60
1061	gi13752754	Homo sapiens	zinc finger 1111	2167	59
1061	AAM79739	Homo sapiens	HYSE- Human protein SEQ ID NO 3385.	2164	54
1062	ABB55706	Homo sapiens	FECH/ Human polypeptide SEQ ID NO 18.	1476	99
1062	AAU38997	Homo sapiens	GEMY Human secreted protein yall 1.	1476	99
1062	AAY17227	Homo sapiens	GEMY Human secreted protein (clone ya1-1).	1476	99
1063	AAB94558	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15328.	593	75
1063	AAM94024	Homo sapiens	HELI- Human stomach cancer expressed polypeptide SEQ ID NO 118.	593	75
1063	gi7379127	Neisseria meningitidis Z2491	phosphoenolpyruvate carboxylase	86	40
1064	gi13752754	Homo sapiens	zinc finger 1111	2345	58
1064	gi10440398	Homo sapiens	FLJ00032 protein	2290	59
1064	AAB95862	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18929.	2085	59
1065	AAO11604	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 25496.	888	100
1065	AAO04671	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 18563.	270	43
1065	AAO01369	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 15261.	202	39

247

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1066	AAE16475	Homo sapiens	OSTE- Human collagen alpha1 (I) protein.	75	38
1066	ABB90764	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 261.	75	38
1066	ABB09625	Homo sapiens	OSTE- Amino acid sequence of human collagen type I alpha1.	75	38
1067	AAG65158	Homo sapiens	BIOU- Human negative regulator 21 of programmed cell death.	710	99
1068	AAB95237	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17384.	300	30
1068	ABB05725	Homo sapiens	GEHU- Human nucleic acid management protein clone tes3_31a10.	298	30
1068	AAB93129	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12017.	298	30
1069	AAB95546	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18167.	2482	99
1069	AAM79264	Homo sapiens	HYSE- Human protein SEQ ID NO 1926.	1016	44
1069	gi8308176	Homo sapiens	cingulin	1016	44
1070	AAB47134	Homo sapiens	INCY- CDIFF-15, Incyte ID No. 3478571CD1.	3992	99
1070	AAU82017	Homo sapiens	INCY- Human secreted protein SECP43.	2195	51
1070	gi18676716	Homo sapiens	FLJ00257 protein	2042	54
1071	gi1230657	Saccharomyces cerevisiae	Ssd1p	162	33
1071	gi172697	Saccharomyces cerevisiae	SRK1	162	33
1071	gi172612	Saccharomyces cerevisiae	SSD1 protein	162	33
1072	AAR60127	Homo sapiens	MASI Human nestin protein is useful to identify brain tumours.	1637	91
1072	gi35019	Homo sapiens	nestin	1637	91
1072	AAR27205	Homo sapiens	MASI Human nestin.	1622	90
1073	AAV91424	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 12 SEQ ID NO:145.	994	96
1073	gi18676600	Homo sapiens	FLJ00198 protein	111	31
1073	gi1532071	Zea mays	glycine-rich protein	96	28
1074	AAE11770	Homo sapiens	INCY- Human kinase (PKIN)-4 protein.	2638	98
1074	gi205278	Rattus norvegicus	male germ cell-associated kinase (mak)	2183	80
1074	gi53914	Mus musculus	rck	2150	79
1075	AAG03354	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7435.	373	100
1075	AAM25814	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1329.	69	30
1075	AAM25290	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:805.	69	30
1076	gi7578595	Mus musculus	teashirt 2	4570	89
1076	gi7527470	Mus musculus	zinc finger protein	2311	48
1076	gi19481304	Danio rerio	teashirt-like zinc finger protein	1785	48
1077	gi14193749	Mus musculus	zinc finger 142	165	28
1077	gi1510147	Homo sapiens	similar to Human zinc finger	164	28

248
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			protein(ZNF142)		
1077	gi6978026	Mus musculus	zinc finger protein ZAC1	161	23
1078	AAAY64650	Homo sapiens	GEST Human human homology protein.	951	99
1078	AAAY01635	Homo sapiens	ABBO Human PS214 derived polypeptide.	951	99
1078	AAAY59682	Homo sapiens	GEST Secreted protein 108-009-5-0-A2-FL.	951	99
1079	AAB36595	Homo sapiens	INCY- Human FLEXHT-17 protein sequence SEQ ID NO:17.	1339	100
1079	gi12002058	Homo sapiens	p5326	1339	100
1079	gi19353786	Mus musculus	expressed sequence AI837181	1305	96
1080	AAM39815	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2960.	451	42
1080	gi14161367	Homo sapiens	megakaryoblastic leukemia-1 protein	451	42
1080	gi14041648	Homo sapiens	OTT-MAL	451	42
1081	AAB36589	Homo sapiens	INCY- Human FLEXHT-11 protein sequence SEQ ID NO:11.	742	100
1081	AAG93289	Homo sapiens	NISC- Human protein HP10641.	742	100
1081	AAU28389	Homo sapiens	HYSE- Novel human secretory protein, Seq ID No 746.	742	100
1082	AAB75321	Homo sapiens	ROSE/ Human secreted protein sequence encoded by gene 34 SEQ ID NO:140.	507	100
1082	gi3875721	Caenorhabditis elegans	F11C1.1	147	32
1082	gi11558246	Mus musculus	calsyntenin-1 protein	85	38
1083	gi18089024	Homo sapiens	Similar to three prime repair exonuclease 1	3331	100
1083	gi17227176	Homo sapiens	ATR-interacting protein	3331	100
1083	AAB93956	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14000.	2919	100
1084	gi21410587	Homo sapiens	similar to RIKEN cDNA 2310041H06	818	100
1084	gi6855513	Gallus gallus	syndesmos	549	63
1084	gi18034388	Mus musculus	syndesmos	537	58
1085	AAB94844	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16020.	672	100
1085	AAU15939	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 892.	672	100
1085	gi20306472	Mus musculus	RIKEN cDNA 2400008I04 gene	506	81
1086	gi12659138	Mus musculus	mage-d3	144	31
1086	gi13516488	Mus musculus	MAGE-necdin/trophinin complex (Magphinin)	138	31
1086	gi12697573	Mus musculus	trophinin-2	138	31
1087	gi2695659	Bos taurus	pyruvate dehydrogenase phosphatase regulatory subunit precursor; PDPr	943	80
1087	gi14022085	Mesorhizobium loti	sarcosine dehydrogenase	346	37
1087	gi15075297	Sinorhizobium meliloti	PUTATIVE OXIDOREDUCTASE PROTEIN	344	39
1088	gi10440476	Homo sapiens	FLJ00075 protein	843	98
1088	AAAY35936	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 185.	395	100
1088	AAB64417	Homo sapiens	INCY- Amino acid sequence of human intracellular signalling molecule	360	100

249
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			INTRA49.		
1089	gi538413	Mus musculus	zinc finger protein	3110	94
1089	gi186774	Homo sapiens	zinc finger protein	2479	56
1089	gi2739353	Homo sapiens	ZNF91L	2437	54
1090	AAG79225	Homo sapiens	GEMY Amino acid sequence of a human PSGL-1 binding protein.	1640	99
1090	AAG79120	Homo sapiens	DAUS- Amino acid sequence of IBD1prox protein.	1640	99
1090	gi14719305	Mus musculus	SNX20	1241	77
1091	AAB99495	Homo sapiens	ARBO- Human CLASP-3 protein sequence SEQ ID NO:2.	6773	99
1091	gi14597912	Homo sapiens	human CLASP-3	6773	99
1091	AAB99493	Homo sapiens	ARBO- Preliminary human CLASP-3 protein sequence Fig 1.	6123	99
1092	gi2425111	Dictyostelium discoideum	ZipA	140	20
1092	gi577221	Saccharomyces cerevisiae	Imh1p	125	20
1092	gi10178157	Arabidopsis thaliana	contains similarity to heat shock protein~gene_id:MPO12.20	119	22
1093	AAB63245	Homo sapiens	LUDW- Human breast cancer associated antigen protein sequence SEQ ID NO:607.	1775	100
1093	gi181247	Homo sapiens	cyclin D3	1514	100
1093	gi15079591	Homo sapiens	Similar to cyclin D3	1514	100
1094	gi8926320	Rattus norvegicus	corneal wound healing related protein	3606	95
1094	gi16768554	Drosophila melanogaster	GM08606p	1275	38
1094	gi19528151	Drosophila melanogaster	AT26759p	970	37
1095	AAB93319	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12406.	2776	99
1095	AAB94051	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14218.	2227	100
1095	AAB90712	Homo sapiens	GEMY Human BG219_1 protein sequence SEQ ID 120.	825	98
1096	AAB43434	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:879.	539	83
1096	AAB57205	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1783.	453	81
1096	AAY07039	Homo sapiens	LUDW- Breast cancer associated antigen precursor sequence.	446	84
1097	gi13543594	Homo sapiens	similar to rat myomegalin	865	98
1097	gi20072822	Homo sapiens	Similar to phosphodiesterase 4D interacting protein (myomegalin)	838	94
1097	gi19263586	Homo sapiens	similar to rat myomegalin	835	97
1098	AAM73511	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 33817.	202	100
1098	AAM60831	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 32936.	202	100
1098	gi156262	Caenorhabditis elegans	collagen	93	35

250
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1099	ABB15372	Homo sapiens	HUMA- Human nervous system related polypeptide SEQ ID NO 4029.	563	84
1099	gi15145795	Sus scrofa	basic proline-rich protein	152	32
1099	gi17945382	Drosophila melanogaster	RE17165p	140	31
1100	gi20271166	Mus musculus	fibrous sheath-interacting protein 1	550	59
1100	gi7800648	Streptococcus pneumoniae	PspA	110	25
1100	gi3171906	Homo sapiens	DIA-156 protein	107	25
1101	gi11989980	Homo sapiens	dJ261G23.2.1 (novel protein, isoform 1)	1704	100
1101	AAM49168	Homo sapiens	SHAN- Human keratin 30.36.	1144	98
1101	gi11989979	Homo sapiens	dJ261G23.2.2 (novel protein, isoform 2)	1144	98
1102	AAB94564	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15341.	2297	99
1102	ABB89620	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1996.	617	96
1102	AAU10981	Homo sapiens	BODE- Human aldehyde/ketone reductase 36.	171	36
1103	gi13276571	Homo sapiens	chr10 synaptotagmin	824	98
1103	gi213111	Discopyge ommata	synaptic vesicle protein	374	33
1103	gi1932801	Rattus norvegicus	synaptotagmin X	370	35
1104	gi18848333	Homo sapiens	Similar to RIKEN cDNA 4930453N24 gene	1740	100
1104	gi18043326	Mus musculus	Similar to RIKEN cDNA 4930453N24 gene	1435	83
1104	AAY73389	Homo sapiens	INCY- HTRM clone 4173111 protein sequence.	1023	100
1105	AAW47209	Homo sapiens	UYNY Homo sapiens tubulin-folding cofactor B.	1042	85
1105	gi2343185	Homo sapiens	tubulin folding cofactor B	1042	85
1105	gi14715044	Mus musculus	RIKEN cDNA 2410007D12 gene	960	77
1106	gi16356663	Homo sapiens	erythrocyte membrane protein 4.1N	4359	96
1106	gi3790545	Mus musculus	neuronal protein 4.1	4245	93
1106	gi4587118	Rattus norvegicus	rat brain 4.1(S)	4233	93
1107	gi16356663	Homo sapiens	erythrocyte membrane protein 4.1N	4527	100
1107	gi3790545	Mus musculus	neuronal protein 4.1	4364	95
1107	gi4587118	Rattus norvegicus	rat brain 4.1(S)	4358	95
1108	AAB93309	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12385.	247	25
1108	gi4584539	Arabidopsis thaliana	extensin-like protein	194	36
1108	gi330444	Human herpesvirus 4	nuclear protein EBNA2	185	37
1109	gi10179839	Homo sapiens	Thy-1 co-transcribed protein	761	100
1109	gi21070344	Homo sapiens	GAS2-related protein isoform beta	100	32
1109	gi2944066	Canis familiaris	synapsin I	98	32
1110	gi15144271	Homo sapiens	NYD-SP11	2037	61
1110	AAO14415	Homo sapiens	UYNA- Human testis development specific protein-11, NYD-SP11.	1934	64

251
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1110	gi10178281	Arabidopsis thaliana	katanin p80 subunit-like protein	113	24
1111	gi15144273	Homo sapiens	protein kinase NYD-SP24	1563	100
1111	gi15823644	Homo sapiens	long form	1559	99
1111	gi18158217	Homo sapiens	calcium-response factor CaRF	1559	99
1112	AAG67079	Homo sapiens	BIOW- Human dihydroorotase 23 polypeptide.	1109	98
1112	gi11994733	Arabidopsis thaliana	contains similarity to cell wall-plasma membrane linker protein~gene_id:MKA23.5	112	24
1112	AAB95856	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18916.	105	27
1113	AAB93872	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13771.	1110	100
1113	gi18146803	Mus musculus	neural-restrictive silencer factor nrsl/rest	104	28
1113	gi560496	Cercopithecine herpesvirus 1	glycoprotein G (homologue of HSV-2 US4)	99	31
1114	AAB94049	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14213.	3253	100
1114	gi15080775	Homo sapiens	protein kinase NYD-SP5	1562	99
1114	AAM72771	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 33077.	239	100
1115	gi20072961	Homo sapiens	similar to zinc finger protein from gene of uncertain exon structure; similar to Q99676 (PID:g3025333)	2763	99
1115	gi4159888	Homo sapiens	zinc finger protein from gene of uncertain exon structure; similar to Q99676 (PID:g3025333)	2212	99
1115	AAM68363	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28669.	1947	100
1116	gi11464740	Homo sapiens	gigaxonin	3160	100
1116	AAG74478	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5242.	757	97
1116	AAG76159	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6923.	640	99
1117	AAU15851	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 804.	2303	99
1117	AAU16312	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1265.	1606	97
1117	AAG02054	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6135.	678	95
1118	gi17390199	Homo sapiens	Similar to zinc finger protein 202	1665	99
1118	AAAY29918	Homo sapiens	MYRI- Human CHD1 protein encoded by transcript cDNA2.	744	38
1118	AAAY29917	Homo sapiens	MYRI- Human CHD1 protein encoded by transcript cDNA1.	744	38
1119	AAA27125 aal	Homo sapiens	UNII Human repressor of estrogen repressor activity (REA) cDNA.	1299	89
1119	ABB11913	Homo sapiens	HYSE- Human B-cell receptor associated protein homologue, SEQ ID NO:2283.	1299	89
1119	AAAY94443	Homo sapiens	UNII Human repressor of estrogen	1299	89

252
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			repressor activity (REA) protein.		
1120	AAG93283	Homo sapiens	NISC- Human protein HP10626.	719	100
1120	gi15559443	Homo sapiens	Similar to RIKEN cDNA 2810417J12 gene	719	100
1120	gi19354196	Mus musculus	RIKEN cDNA 2810417J12 gene	500	71
1121	AAB58852	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 560.	511	100
1121	AAM39029	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2174.	75	30
1121	gi6899807	Homo sapiens	zinc finger protein 219	75	30
1122	gi8099348	Homo sapiens	zinc finger protein	1378	52
1122	AAM00913	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 389.	1369	62
1122	AAM78963	Homo sapiens	HYSE- Human protein SEQ ID NO 1625.	1369	62
1123	AAG02740	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6821.	458	98
1123	gi2951763	Sugarcane mosaic virus	capsid protein	78	26
1123	gi1249618	Johnsongrass mosaic virus	coat protein	78	26
1124	gi19483971	Mus musculus	RIKEN cDNA 3230401I01 gene	2313	85
1124	gi2576348	Homo sapiens	Glutamyl tRNA synthetase	1520	100
1124	gi21064393	Drosophila melanogaster	RE18828p	1115	48
1125	AAB82047	Homo sapiens	IGAK- Human mast cell surface antigen.	2141	100
1125	gi12654783	Homo sapiens	Similar to loss of heterozygosity, 11, chromosomal region 2, gene A	2141	100
1125	gi2190974	Homo sapiens	breast cancer suppressor candidate 1	338	100
1126	AAB95305	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17544.	3327	99
1126	AAB92704	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11106.	2993	99
1126	AAB94137	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14404.	2944	100
1127	gi14861050	Ornithorhynchus anatinus	sulfotransferase SULT1A	73	39
1127	gi5081624	Emericella nidulans	molybdopterin synthase large subunit CnxH	70	37
1127	gi21292708 gb EAA048 53.1	Anopheles gambiae str. PEST	agCP4275	80	29
1128	gi5457306	Homo sapiens	dJ1065O2.3 (paired box gene 1)	2358	100
1128	gi9501358	Mus musculus	Pax1 protein	1681	90
1128	gi12658965	Mus musculus	PAX1	1681	90
1129	gi18642466	Homo sapiens	EMSY protein	1915	99
1129	AAW82571	Homo sapiens	CANC- Human BBP1 DNA.	1825	95
1129	AAM93627	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3466.	623	100
1130	AAV10834	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	1603	99
1130	AAB93801	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13573.	1597	99

253
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1130	gi17946001	Drosophila melanogaster	RE50056p	254	31
1131	AAAY10834	Homo sapiens	HUMA- Amino acid sequence of a human secreted protein.	1565	100
1131	AAB93801	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13573.	1559	99
1131	gi17946001	Drosophila melanogaster	RE50056p	253	31
1132	gi1504002	Homo sapiens	similar to a human major CRK-binding protein DOCK180.	9481	99
1132	gi13195147	Mus musculus	HCH	9027	94
1132	gi1339910	Homo sapiens	DOCK180 protein	5541	62
1134	gi3600087	Dictyostelium discoideum	cytosolic regulator pianissimo	93	34
1134	gi3411250	Dictyostelium discoideum	developmental protein	93	34
1134	gi15011308	Human immunodeficiency virus type 1	envelope glycoprotein	82	27
1135	gi13543686	Homo sapiens	Similar to RIKEN cDNA 4931428F02 gene	1483	98
1135	AAAY07081	Homo sapiens	LUDW- Renal cancer associated antigen precursor sequence.	255	40
1135	AAE13761	Homo sapiens	CORI- Human lung tumour-specific protein LT86-14.	254	37
1136	gi14250138	Homo sapiens	Similar to RIKEN cDNA 5730421E18 gene	1377	100
1136	gi18605826	Mus musculus	Similar to RIKEN cDNA 5730421E18 gene	887	91
1136	AAU20419	Homo sapiens	HUMA- Human secreted protein, Seq ID No 411.	850	95
1137	AAU20488	Homo sapiens	HUMA- Human secreted protein, Seq ID No 480.	643	83
1137	AAM93522	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3253.	643	83
1137	AAAY57916	Homo sapiens	INCY- Human transmembrane protein HTMPN-40.	643	83
1138	gi6934248	Homo sapiens	tropomodulin 4	1536	89
1138	gi12744762	Homo sapiens	muscle tropomodulin	1536	89
1138	gi6013189	Homo sapiens	Sk-tropomodulin	1530	89
1139	AAB43900	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1345.	516	89
1139	gi1666702	Mus musculus	ribosomal protein	516	89
1139	gi206732	Rattus norvegicus	ribosomal protein L36a	516	89
1140	AAG00664	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4745.	619	100
1140	AAG00663	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4744.	489	100
1140	ABB04721	Homo sapiens	SHAN- Human PP2464 protein SEQ ID NO:17.	89	26
1141	gi14250752	Homo sapiens	Similar to hect domain and RLD 2	890	100
1141	gi16648386	Drosophila melanogaster	LD39062p	347	31
1141	gi4079809	Homo sapiens	HERC2	288	32

254
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1142	gi15277241	Homo sapiens	Testis expressed gene	485	90
1142	gi3176438	Homo sapiens	HCG V	485	90
1142	gi11322969	Homo sapiens	protein phosphatase 1, regulatory (inhibitor) subunit 11	485	90
1143	AAB64461	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 40 SEQ ID NO:99.	272	100
1143	AAB75511	Homo sapiens	ROSE/ Human secreted protein sequence encoded by gene 6 SEQ ID NO:65.	272	100
1143	AAB51649	Homo sapiens	ROSE/ Human secreted protein sequence encoded by gene 30 SEQ ID NO:89.	272	100
1144	AAB43827	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1272.	515	87
1144	gi17944934	Drosophila melanogaster	RE02292p	400	36
1144	AAB64374	Homo sapiens	INCY- Amino acid sequence of human intracellular signalling molecule INTRA6.	175	29
1145	gi14549207	Homo sapiens	URAX1	3118	100
1145	gi17907795	Homo sapiens	TGF-beta induced apoptosis protein 3	3118	100
1145	AAM51815	Homo sapiens	BADI Human apoptase L100.	3117	99
1146	gi15981951	Yersinia pestis	L-seryl-tRNA selenium transferase	67	47
1147	AAG74155	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4919.	1118	98
1147	AAB53429	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:969.	1118	98
1147	gi19353875	Mus musculus	RIKEN cDNA 2210417D09 gene	905	73
1148	gi4050093	Mus musculus	NG28	2465	75
1148	AAM38700	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 1845.	753	42
1148	AAM38699	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 1844.	749	45
1149	gi52977	Mus musculus	modifier 3 (M33)	2181	82
1149	gi2266988	Mus musculus	M33 polycomb-like protein	2161	82
1149	gi3860185	Xenopus laevis	Polycomb homolog Pc1	950	45
1150	gi15213542	Homo sapiens	NSD1	1618	100
1150	gi16751269	Homo sapiens	androgen receptor associated coregulator 267-b	1614	100
1150	gi3329465	Mus musculus	NSD1 protein	1455	90
1151	AAU75614	Homo sapiens	BIOW- Human RS3 protein 18.	922	100
1151	gi12005728	Homo sapiens	GL012	922	100
1151	AAU75615	Homo sapiens	BIOW- Human RS3 protein 18, N terminal peptide.	78	100
1152	AAU82017	Homo sapiens	INCY- Human secreted protein SECP43.	2464	87
1152	gi18676716	Homo sapiens	FLJ00257 protein	2356	84
1152	AAM00753	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 116.	1640	91
1153	gi6572156	Homo sapiens	dJ1014D13.3 (novel protein)	1783	100
1153	gi6572155	Homo sapiens	dJ1014D13.2 (novel protein similar to ACTN3 (actinin, alpha 3))	1023	100
1153	gi18676484	Homo sapiens	FLJ00139 protein	704	35
1154	AAG03284	Homo sapiens	GEST Human secreted protein, SEQ	364	100

255

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			ID NO: 7365.		
1154	gi4778	Saccharomyces cerevisiae	Uso1 protein	119	22
1154	gi2853301	Homo sapiens	mucin	107	20
1155	AAM25280	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:795.	1908	100
1155	gi19338692	Mus musculus	N-acetylglutamate synthase	1780	92
1155	AAB58405	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 743.	814	99
1156	AAW21949	Homo sapiens	NEWE- E6-binding protein E6-BPSD7.	1678	100
1156	gi469885	Homo sapiens	EF-hand protein	1678	100
1156	gi13436152	Homo sapiens	reticulocalbin 2, EF-hand calcium binding domain	1678	100
1157	gi18676654	Homo sapiens	FLJ00226 protein	5094	99
1157	gi1293893	Mus musculus	leucine zipper protein 1	4324	80
1157	gi19880160	Mus musculus	leucine zipper motif-containing protein	4295	80
1158	AAB94413	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15003.	1044	94
1158	gi12804385	Homo sapiens	HT002 protein; hypertension-related calcium-regulated gene	1044	94
1158	AAB93000	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11737.	1038	93
1159	gi3387790	Homo sapiens	PIR1	1407	80
1159	gi12653157	Homo sapiens	dual specificity phosphatase 11 (RNA/RNP complex 1-interacting)	1399	79
1159	AAG01376	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5457.	542	100
1160	AAM88829	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:16422.	1017	94
1160	gi5901688	Mus musculus	GRIN1	251	24
1160	gi5650774	Gallus gallus	interacting protein 6	186	32
1161	AAM67381	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27687.	814	100
1161	AAM54995	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 27100.	814	100
1161	AAU16451	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1404.	630	100
1162	gi13661556	Homo sapiens	extracellular glycoprotein EMILIN-2 precursor	1410	98
1162	gi21434745	Mus musculus	basilin	1099	71
1162	AAM99923	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 39.	1096	93
1163	AAM75827	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36133.	648	100
1163	AAM63016	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 35121.	648	100
1163	gi46363	Rhodospirillum rubrum	URF2	88	25
1164	AAU76957	Homo sapiens	CYTO- Novel human kinesin motor	4649	100

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			protein, HsKip3d.		
1164	gi21104476	Homo sapiens	OK/SW-CL.108	4637	99
1164	AAU76958	Homo sapiens	CYTO- Novel human kinesin motor protein, HsKip3d motor domain.	1835	100
1165	AAE04764	Homo sapiens	INCY- Human vesicle trafficking protein-7 (VETRP-7) protein.	2533	100
1165	gi15186738	Mus musculus	Tac2-N	2226	87
1165	AAU19746	Homo sapiens	HUMA- Human novel extracellular matrix protein, Seq ID No 396.	998	100
1166	gi2760351	Girardia tigrina	myosin heavy chain	197	23
1166	gi3986194	Dugesia japonica	myosin heavy chain	197	22
1166	gi18652658	Schmidtea mediterranea	myosin heavy chain A	179	22
1167	AAV56021	Homo sapiens	VLAA- Human CD40 receptor interacting protein 4C4.	2121	99
1167	AAO12585	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26477.	548	96
1167	AAM40573	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5504.	341	22
1168	gi20126671	Rattus norvegicus	socius	1971	78
1168	gi18044095	Mus musculus	Similar to RIKEN cDNA 4930506L07 gene	1199	82
1168	gi13195163	Rattus norvegicus	stromal antigen 3	143	43
1169	gi18072031	Homo sapiens	zinc finger protein 328	234	58
1169	gi14456629	Homo sapiens	dJ54B20.2 (novel KRAB box containing C2H2 type zinc finger protein)	220	52
1169	AAO11585	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 25477.	219	54
1170	gi20381091	Homo sapiens	similar to rab11-binding protein	85	26
1170	gi257041	Zea mays	hydroxyproline-rich glycoprotein; HRGP	83	23
1170	gi4007865	Zea mays	Hydroxyproline-rich Glycoprotein (HRGP)	83	23
1171	gi11415020	Rattus norvegicus	calcium channel alpha-1-I subunit	70	28
1171	gi18593484 [refXP_092855.1]	Homo sapiens	similar to interspersed repeat antigen	73	32
1171	gi20904051 [refXP_139476.1]	Mus musculus	similar to Voltage-dependent T-type calcium channel alpha-1I subunit (CAVT.3)	71	30
1172	AAB94634	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15508.	1144	47
1172	AAW52187	Homo sapiens	TEXA Human BRCA1-associated protein (hBRAP) primary sequence.	1144	47
1172	gi10442700	Homo sapiens	zinc-finger protein ZBRK1	1144	47
1173	gi21410587	Homo sapiens	similar to RIKEN cDNA 2310041H06	528	100
1173	gi6855513	Gallus gallus	syndesmos	414	64
1173	gi19343886	Mus musculus	Similar to RIKEN cDNA 1110001K21 gene	407	60
1174	AAU18886	Homo sapiens	HUMA- Novel prostate gland antigen, Seq ID No 185.	370	97

257
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1174	AAM96038	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4696.	370	97
1174	gi13310782	Mus musculus	myoneurin	231	32
1175	gi21105480	Rattus norvegicus	embryo-related protein	1356	93
1175	gi15292203	Drosophila melanogaster	LD42056p	963	64
1175	AAB58899	Homo sapiens	HUMA- Breast and ovarian cancer associated antigen protein sequence SEQ ID 607.	843	99
1176	gi20380058	Homo sapiens	Similar to PRAM-1 protein	2636	99
1176	gi11558109	Homo sapiens	adaptor molecule-1	2572	90
1176	gi13938357	Homo sapiens	Similar to RIKEN cDNA 0610030H11 gene	1435	88
1177	AAB93711	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13305.	3044	100
1177	gi21427588	Homo sapiens	Athabaskan SCID transcript variant 4	3034	99
1177	gi21427586	Homo sapiens	Athabaskan SCID transcript variant 3	3034	99
1178	gi18676446	Homo sapiens	FLJ00120 protein	2006	100
1178	gi17046299	Homo sapiens	CARD-containing MAGUK protein CARMA1	2006	100
1178	AAU01207	Homo sapiens	MILL- Human caspase recruitment domain, CARD-11 polypeptide.	1999	99
1179	AAW37863	Homo sapiens	SAGA Amino acid sequence of the human type-1 membrane protein.	691	100
1179	gi4583677	Homo sapiens	p24B protein	691	100
1179	gi18490109	Homo sapiens	integral type I protein	691	100
1180	AAB47327	Homo sapiens	CURA- FCTR4.	3068	99
1180	AAE18118	Homo sapiens	MILL- Human guanine nucleotide exchange factor (GEF) 32529 protein.	3058	99
1180	ABB06051	Homo sapiens	COMP- Human NS protein sequence SEQ ID NO:143.	2962	97
1181	gi10636484	Homo sapiens	polyglutamine-containing protein	4193	100
1181	AAG81247	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:12.	1180	100
1181	AAM78881	Homo sapiens	HYSE- Human protein SEQ ID NO 1543.	1072	44
1182	AAM40320	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3465.	103	24
1182	AAM40319	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3464.	103	24
1182	AAB93849	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13697.	96	25
1183	AAB74944	Homo sapiens	YAMA Human ADAM type metal protease MDT51 protein SEQ ID NO:1.	109	27
1183	AAB86949	Homo sapiens	HOFF Human metalloprotease MPTS-19 protein.	109	27
1183	gi89777890	Homo sapiens	ADAMTS7, alternatively spliced product	109	27
1184	AAB94379	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14928.	3306	99
1184	gi20072682	Mus musculus	actin-related protein 8 homolog (S. cerevisiae)	3241	97
1184	AAG66735	Homo sapiens	UYFU- Actin 41.	1960	99

258
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1185	gi17066106	Homo sapiens	Novex-3 Titin Isoform	10131	99
1185	AAE04360	Homo sapiens	INCY- Human kinase (PKIN)-1.	2456	100
1185	AAU17981	Homo sapiens	HUMA- Human immunoglobulin polypeptide SEQ ID No 126.	1345	99
1186	gi13276231	Homo sapiens	FYVE and coiled-coil domain containing 1	7470	100
1186	gi18250726	Mus musculus	FYVE and coiled-coil domain containing 1	5550	75
1186	gi13938195	Homo sapiens	FYVE and coiled-coil domain containing 1	1271	98
1187	AAB94590	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15398.	1290	100
1187	gi13938195	Homo sapiens	FYVE and coiled-coil domain containing 1	1263	99
1187	gi18250726	Mus musculus	FYVE and coiled-coil domain containing 1	847	70
1188	gi1235527	Homo sapiens	myoglobin	809	100
1188	gi15778933	Homo sapiens	Similar to myoglobin	809	100
1188	gi386872	Homo sapiens	myoglobin	806	99
1189	gi14587078	Danio rerio	Unc119c	851	69
1189	gi14587076	Danio rerio	Unc119b	780	76
1189	gi4103940	Mus musculus	UNC-119	701	60
1190	gi21303413	Homo sapiens	adenylosuccinate synthetase isozyme	2399	100
1190	gi404751	Mus musculus	adenylosuccinate synthetase	2353	96
1190	gi415849	Homo sapiens	adenylosuccinate synthetase	1861	77
1191	gi10834696	Homo sapiens	PP4189	744	98
1191	gi15081791	Arabidopsis thaliana	At1g09280/T12M4_1	510	36
1191	gi13702647	Staphylococcus aureus subsp. aureus N315	conserved hypotehtical protein	438	33
1192	gi2970646	Mus musculus	Xin	3605	52
1192	gi2970644	Gallus gallus	Xin	1777	50
1192	AAO05743	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 19635.	629	93
1193	gi19343859	Mus musculus	Similar to vesicle amine transport protein 1	1826	87
1193	gi602278	Homo sapiens	similar to Pacific ray VAT1 protein, Swiss-Prot Accession Number P19333	1579	99
1193	gi12804921	Homo sapiens	Similar to membrane protein of cholinergic synaptic vesicles	1556	100
1194	AAB99541	Homo sapiens	ARBO- Human CLASP-7 protein sequence SEQ ID NO:2.	4269	99
1194	gi14598037	Homo sapiens	human CLASP-7	4269	99
1194	AAB99494	Homo sapiens	ARBO- Human CLASP-3 protein sequence Fig 4.	3153	72
1195	gi11493417	Homo sapiens	PRO1292	244	100
1196	AAB60467	Homo sapiens	INCY- Human cell cycle and proliferation protein CCYPR-15, SEQ ID NO:15.	972	100
1196	AAG93315	Homo sapiens	NISC- Human protein HP10456.	972	100
1196	AAM41305	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6236.	972	100
1197	AAU78233	Homo sapiens	PFIZ Human VDUP-1 protein.	1631	99
1197	gi11559490	Homo sapiens	VDUP1	1631	99

259

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1197	gi7717464	Homo sapiens	brain-expressed HHCPA78 homolog VDUP1	1631	99
1198	gi14906268	Homo sapiens	Pur-beta	1638	100
1198	gi2460119	Mus musculus	vascular actin single-stranded DNA-binding factor 2 p44 component; purine-rich single-stranded DNA-binding protein beta; PurB beta	1573	94
1198	AAV31720_aal	Homo sapiens	MOUN Nucleotide sequence of the PUR-alpha gene.	990	71
1199	gi17945201	Drosophila melanogaster	RE12402p	487	30
1199	gi17945128	Drosophila melanogaster	RE08932p	459	28
1199	gi5668787	Arabidopsis thaliana	F21H2.12	87	28
1200	AAM88553	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:16146.	623	98
1200	gi17945760	Drosophila melanogaster	RE33302p	559	45
1200	gi1039447	Saccharomyces cerevisiae	Lpb1p	313	30
1201	gi16610187	Homo sapiens	NSE2 protein	1663	100
1201	gi20070972	Homo sapiens	similar to NSE1 protein	618	46
1201	gi16505737	Homo sapiens	NSE1 protein	618	46
1202	gi19353084	Homo sapiens	Similar to RIKEN cDNA 1700026A16 gene	3951	100
1202	gi13561413	Homo sapiens	F-BOX domain protein	2076	99
1202	gi6561831	Homo sapiens	muscle disease-related protein	1054	34
1203	gi219980	Homo sapiens	paraneoplastic cerebellar degeneration-associated antigen	2243	100
1203	gi241777	Homo sapiens	autoantigen recognized by an anti-neuronal cell antibody	2243	100
1203	gi17028383	Homo sapiens	Similar to cerebellar degeneration-related 2	2243	100
1204	gi11870006	Rattus norvegicus	nuclear matrix transcription factor	737	81
1204	gi18700313	Homo sapiens	nuclear matrix transcription factor4	720	80
1204	gi6729087	Rattus norvegicus	Cas-associated zinc finger protein	716	80
1205	AAB94801	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15933.	466	100
1205	AAO12222	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26114.	466	100
1205	gi13537206	Homo sapiens	Mel18 and Bmi1 like ring finger	466	100
1206	gi18478557	Rattus norvegicus	paraneoplastic onconeural protein MA1	181	48
1206	AAB74701	Homo sapiens	INCY- Human membrane associated protein MEMAP-7.	178	48
1206	gi14030861	Homo sapiens	paraneoplastic neuronal antigen MA1	178	48
1207	AAAY72162	Homo sapiens	BAUG/ Human RNA metabolism protein (RMEP-2).	2791	90
1207	AAB56895	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1473.	2759	97
1207	AAW74802	Homo sapiens	HUMA- Human secreted protein	2759	97

260

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			encoded by gene 73 clone HSQEL25.		
1208	gi21262188	Homo sapiens	CTCL tumor antigen L14-2	926	95
1208	gi15292291	Drosophila melanogaster	LD45682p	171	26
1208	gi20087019	Plasmodium falciparum	maebl	157	28
1209	gi15991574	Homo sapiens	eferin	97	29
1209	gi21428632	Drosophila melanogaster	LP07116p	94	28
1209	gi4176552	Schizosaccharomyces pombe	protein kinase bub1p	84	26
1210	gi18257326	Mus musculus	RIKEN cDNA 2410002O22 gene	2089	96
1210	AAB95495	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18037.	1817	100
1210	gi15080536	Homo sapiens	Similar to RIKEN cDNA 2410002O22 gene	1026	100
1211	gi18257326	Mus musculus	RIKEN cDNA 2410002O22 gene	2046	94
1211	AAB95495	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18037.	1774	98
1211	gi15080536	Homo sapiens	Similar to RIKEN cDNA 2410002O22 gene	983	97
1212	AAU15918	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 871.	2725	99
1212	ABB56422	Homo sapiens	SHAN- Human cancer suppressor protein PP902.	1774	100
1212	AAU16370	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1323.	1400	98
1213	gi19353360	Mus musculus	RIKEN cDNA 2010000G05 gene	473	98
1213	gi268	Bos taurus	cytochrome c oxidase subunit VIb (AA 1-86)	445	90
1213	AAO13877	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 27769.	423	83
1214	gi13279335	Homo sapiens	pleckstrin homology, Sec7 and coiled/coil domains 2 (cytohesin-2)	1924	93
1214	gi1575766	Homo sapiens	cytohesin-2	1924	93
1214	gi3885503	Mus musculus	cytohesin-2	1913	92
1215	gi5138916	Homo sapiens	PTD012	1116	99
1215	gi16182641	Drosophila melanogaster	GH07301p	721	47
1215	gi19527625	Drosophila melanogaster	RE72485p	665	44
1216	AAM49106	Homo sapiens	SHAN- Human beta transduction factor 14.	776	100
1216	AAG66887	Homo sapiens	SHAN- Human zinc finger protein 17.	230	37
1216	AAG66795	Homo sapiens	BIOW- Human ATP-dependent serine protease 16.	152	30
1217	gi20380145	Homo sapiens	FLJ00166 protein	1753	100
1217	gi15209353	Caenorhabditis elegans	Y39B6A.1	186	28
1217	gi8096269	Nicotiana tabacum	KED	153	23
1218	ABB89403	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 1779.	792	96
1218	AAB56606	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1184.	792	96

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1218	AAG01959	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6040.	567	99
1219	gi1184949	Homo sapiens	phosphotyrosine independent ligand for the Lck SH2 domain p62	1821	94
1219	gi17512269	Homo sapiens	sequestosome 1	1821	94
1219	gi12804857	Homo sapiens	Similar to sequestosome 1	1821	94
1220	gi10440530	Homo sapiens	FLJ00115 protein	4008	99
1220	gi5917730	Homo sapiens	F-box protein Lilina	2467	99
1220	AAB93188	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12140.	1050	46
1221	AAM95333	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 3991.	238	38
1221	gi14039845	Homo sapiens	testes development-related NYD-SP18	142	29
1221	AAM95425	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4083.	123	30
1222	gi9802536	Arabidopsis thaliana	F17L21.25	97	23
1222	gi16226498	Arabidopsis thaliana	At1g27460/F17L21_26	97	23
1222	gi5002645	Homo sapiens	IF2 protein	94	22
1223	gi16197827	Drosophila melanogaster	GH09755p	605	52
1223	AAAY68783	Homo sapiens	INCY- Amino acid sequence of a human phosphorylation effector PHSP-15.	538	41
1223	gi3116214	Homo sapiens	SH3 binding protein	538	41
1224	gi18087408	Homo sapiens	HMG-box transcription factor BBXa	4731	99
1224	gi18087406	Homo sapiens	HMG-box transcription factor BBX	4731	99
1224	gi18418655	Homo sapiens	HMG-box containing protein	4690	96
1225	AAB21042	Homo sapiens	INCY- Human nucleic acid-binding protein, NuABP-46.	2747	96
1225	ABB11193	Homo sapiens	HYSE- Human Zn finger protein homologue, SEQ ID NO:1563.	2142	99
1225	gi18204788	Mus musculus	RIKEN cDNA 2810411K16 gene	1286	71
1226	gi2217970	Homo sapiens	p40	1231	66
1226	gi12653463	Homo sapiens	Rab9 effector p40	1228	66
1226	AAM25756	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1271.	1070	61
1227	gi2217970	Homo sapiens	p40	1597	81
1227	gi12653463	Homo sapiens	Rab9 effector p40	1588	81
1227	AAM25756	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1271.	1436	75
1228	gi6572379	Homo sapiens	dJ579N16.2 (SET binding factor 1)	2962	62
1228	gi3015538	Homo sapiens	nuclear dual-specificity phosphatase	2955	62
1228	gi15292603	Drosophila melanogaster	SD10541p	2006	45
1229	ABB50213	Homo sapiens	INCY- Human transcription factor TRFX-64.	2087	100
1229	gi10334466	Homo sapiens	dJ63P18.1 (novel zinc finger protein)	2063	100
1229	AAM78755	Homo sapiens	HYSE- Human protein SEQ ID NO 1417.	937	43
1230	gi19908264	Homo sapiens	N-RAP-like protein	2766	94
1230	gi2351568	Mus musculus	N-RAP	1702	56
1230	AAU21865	Homo sapiens	HUMA- Human cardiovascular system	1136	98

262
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			antigen polypeptide SEQ ID No 639.		
1231	ABB04616	Homo sapiens	BODA- Human TPRs structural domain protein SEQ ID NO:2.	2257	99
1231	AAM93700	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3623.	2257	99
1231	AAW82500	Homo sapiens	USSH Human OGT protein.	129	24
1232	AAM40418	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3563.	6480	100
1232	AAM42202	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 7133.	6296	100
1232	AAW01825	Homo sapiens	CLLT Human metalloproteinase.	107	34
1233	gi12655239	Homo sapiens	bA472E5.2.1 (novel protein, isoform 1)	1179	100
1233	gi17512323	Mus musculus	Similar to RIKEN cDNA 2610016C23 gene	1135	66
1233	gi12655240	Homo sapiens	bA472E5.2.2 (novel protein, isoform 2)	994	98
1234	AAG78904	Homo sapiens	BIOW- Human ribosome S7 protein 16.	751	100
1234	gi12005904	Homo sapiens	AD036	751	100
1234	AAG78905	Homo sapiens	BIOW- Human ribosome S7 protein 16 peptide fragment.	77	100
1235	gi19068693	Encephalitozoon cuniculi	similarity to RAD4 protein	211	35
1235	gi20073005	Mus musculus	Similar to topoisomerase (DNA) II binding protein	194	27
1235	AAV03183	Homo sapiens	TSUR/ Topoisomerase II binding protein 2.	192	26
1236	AAG63832	Homo sapiens	GENE- Amino acid sequence of human cardiomyopathy associated protein (CAP).	1869	100
1236	AAG93278	Homo sapiens	NISC- Human protein HP03115.	1869	100
1236	gi9945010	Mus musculus	RING-finger protein MURF	1776	95
1237	gi11908000	Homo sapiens	BCL-6 corepressor short isoform	5272	100
1237	gi11907998	Homo sapiens	BCL-6 corepressor	5246	99
1237	AAM96154	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4812.	290	96
1238	gi12655378	Homo sapiens	oxodicarboxylate carrier	1552	100
1238	gi12655649	Rattus norvegicus	mitochondrial oxodicarboxylate carrier	1297	82
1238	gi18447084	Drosophila melanogaster	AT23224p	581	41
1239	gi1526421	Mus musculus	KAP3B	115	21
1239	gi1526419	Mus musculus	KAP3A	115	21
1239	AAB07973	Homo sapiens	UTOR A human neural plakophilin related armidillo protein.	102	39
1240	AAV01070	Homo sapiens	SUME Human I(3)mbt protein sequence.	1461	48
1240	gi3811111	Homo sapiens	I(3)mbt protein homolog	1461	48
1240	gi11323323	Homo sapiens	dJ138B7.3.1 (continued from dJ862K6.1 in Em:AL031681)	1447	51
1241	AAG00343	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4424.	277	94
1241	AAM89138	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:16731.	260	97
1241	gi121296312	Anopheles	ebiP3199	441	28

263
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
	[gb]EAA084 57.1	gambiae str. PEST			
1242	gi20135645	Homo sapiens	SVAP1 protein	2376	99
1242	gi8778388	Arabidopsis thaliana	F16A14.10	93	28
1242	gi500838	Saccharomyces cerevisiae	Yhr080cp	92	25
1243	gi14330448	Homo sapiens	zinc finger protein RINZF	4282	95
1243	gi4557143	Rattus norvegicus	zinc finger protein RIN ZF	2792	78
1243	AAB94280	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14711.	1343	100
1244	gi17389709	Homo sapiens	Similar to RIKEN cDNA 1110001A07 gene	608	100
1244	AAG03897	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7978.	551	91
1244	AAB93244	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12253.	77	29
1245	gi21040415	Homo sapiens	TruB pseudouridine synthase-like protein 1	1747	100
1245	AAAY35964	Homo sapiens	GEST Extended human secreted protein sequence, SEQ ID NO. 213.	839	99
1245	gi9951004	Pseudomonas aeruginosa	tRNA pseudouridine 55 synthase	373	39
1246	gi15341790	Homo sapiens	Similar to RIKEN cDNA 2900009I07 gene	751	60
1246	gi20306430	Mus musculus	RIKEN cDNA 2900009I07 gene	746	59
1246	AAG74211	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4975.	597	96
1247	AAM80283	Homo sapiens	HYSE- Human protein SEQ ID NO 3929.	1202	58
1247	gi6409345	Homo sapiens	zinc finger protein ZNF180	1196	60
1247	gi8050899	Homo sapiens	ZNF180	1196	60
1248	ABB90394	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2770.	1483	99
1248	gi15214665	Homo sapiens	Similar to RIKEN cDNA 2310061O04 gene	824	100
1248	AAB56403	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:981.	813	98
1249	AAB60476	Homo sapiens	INCY- Human cell cycle and proliferation protein CCYPR-24, SEQ ID NO:24.	1592	100
1249	gi14603190	Homo sapiens	MAGEF1 protein	1592	100
1249	gi11096137	Homo sapiens	MAGEF1	1592	100
1250	gi14043262	Homo sapiens	Similar to RIKEN cDNA 1500026B10 gene	1076	100
1250	gi20073258	Mus musculus	RIKEN cDNA 1100001D19 gene	758	73
1250	gi17646146	Homo sapiens	B lymphocyte activation-related protein	462	47
1251	AAE06142	Homo sapiens	HUMA- Human gene 7 encoded secreted protein fragment, SEQ ID NO:204.	566	98
1251	AAE06047	Homo sapiens	HUMA- Human gene 7 encoded secreted protein HBJFJ26, SEQ ID NO:109.	566	98

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1251	AAV87165	Homo sapiens	HUMA- Human secreted protein sequence SEQ ID NO:204.	566	98
1252	gi2477513	Homo sapiens	F25965_3	2718	87
1252	AAE13842	Homo sapiens	CORI- Human lung tumour-specific protein 20129.	540	33
1252	gi15145797	Sus scrofa	basic proline-rich protein	330	31
1253	gi15139362	Mus musculus	aczonin	6067	96
1253	gi7493836	Rattus norvegicus	multidomain presynaptic cytomatrix protein Piccolo	6043	95
1253	gi6433844	Gallus gallus	aczonin	5501	87
1254	gi15139362	Mus musculus	aczonin	6006	95
1254	gi7493836	Rattus norvegicus	multidomain presynaptic cytomatrix protein Piccolo	5982	95
1254	gi6433844	Gallus gallus	aczonin	5440	86
1255	AAV39467	Homo sapiens	ABGE- Heterogenous ribonuclear protein K.	2307	94
1255	gi409389	Rattus norvegicus	dC-stretch binding protein (CSBP)	2307	94
1255	gi241478	Homo sapiens	heterogeneous nuclear ribonucleoprotein complex K; hnRNP K	2307	94
1256	AAW47589	Homo sapiens	BOEF T-cell receptor beta-chain.	537	91
1256	gi1552495	Homo sapiens	V segment translation product	534	99
1256	gi3821867	Homo sapiens	T cell receptor beta chain variable region	532	90
1257	AAM94177	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 2835.	370	100
1257	gi8096269	Nicotiana tabacum	KED	329	22
1257	gi7549210	Babesia bigemina	200 kDa antigen p200	298	25
1258	AAG93327	Homo sapiens	NISC- Human protein HP10502.	1438	100
1258	AAB43628	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1073.	1101	99
1258	gi20379943	Mus musculus	Similar to RIKEN cDNA 4930553M18 gene	612	59
1259	gi18676506	Homo sapiens	FLJ00150 protein	8766	99
1259	gi3435244	Homo sapiens	centriole associated protein CEP110	4915	99
1259	gi18070853	Homo sapiens	bA165P4.2 (centrosomal protein 1)	4887	98
1260	ABB06784	Homo sapiens	SHAN- Human cancer cell growth inhibiting protein PP4762 SEQ ID NO:20.	205	39
1260	gi10834714	Homo sapiens	PP4762	205	39
1260	gi3878051	Caenorhabditis elegans	H21P03.2	169	31
1261	ABB06133	Homo sapiens	COMP- Human NS protein sequence SEQ ID NO:225.	1626	93
1261	AAM25414	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:929.	874	97
1261	AAV07040	Homo sapiens	LUDW- Breast cancer associated antigen precursor sequence.	856	39
1262	gi14550457	Homo sapiens	pellino (Drosophila) homolog 2	2290	100
1262	gi10242353	Homo sapiens	pellino 2	2290	100
1262	gi10242357	Mus musculus	pellino 2	2110	92
1263	gi18478648	Homo sapiens	EKN1	1790	98

265
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1263	gi20071434	Mus musculus	Similar to EKN1	1389	75
1263	gi19527855	Drosophila melanogaster	AT10515p	165	25
1264	gi632549	Petromyzon marinus	NF-180	220	24
1264	gi20452161	Canis familiaris	retinitis pigmentosa GTPase regulator	220	28
1264	gi9837379	Homo sapiens	retinitis pigmentosa GTPase regulator	210	28
1265	gi14517637	Homo sapiens	RGPR-p117	87	30
1265	gi12313999	Homo sapiens	dJ885L7.3.1 (opioid growth factor receptor (7-60 protein), isoform 1)	81	31
1265	AAV92809	Homo sapiens	PENN- Human opioid growth factor receptor spliced version 8.	78	28
1266	gi18204361	Homo sapiens	PNAS-127 protein	380	100
1266	gi12751100	Homo sapiens	PNAS-127	380	100
1266	gi14210498 [ref]NP_115879.1]	Homo sapiens	PNAS-127 protein	380	100
1267	ABB50177	Homo sapiens	INCY- Human transcription factor TRFX-28.	1876	100
1267	gi11611473	Mus musculus	Deltex3	1858	98
1267	AAU16308	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1261.	1302	96
1268	AAB95168	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17221.	2667	99
1268	gi13278316	Mus musculus	Similar to RIKEN cDNA 2610507L03 gene	2174	70
1268	AAG75001	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5765.	776	95
1269	gi6016842	Mus musculus	nuclear protein ZAP	3147	61
1269	AAV72168	Homo sapiens	BAUG/ Human RNA metabolism protein (RMEP-8).	2798	100
1269	AAM75582	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 35888.	788	100
1270	AAB93344	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12459.	92	95
1270	AAO13507	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 27399.	92	95
1270	AAO13502	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 27394.	92	95
1271	AAG81396	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:310.	629	100
1271	AAG00714	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4795.	527	99
1271	gi18447274	Drosophila melanogaster	HL01254p	287	45
1272	gi15292097	Drosophila melanogaster	LD38226p	430	27
1272	gi2414626	Schizosaccharomyces pombe	ubiquitin regulatory domain (UBX) protein	404	31
1272	gi7527718	Arabidopsis thaliana	T5E21.7	336	37
1273	gi4263747	Homo sapiens	general transcription factor 2I	730	79
1273	gi2415382	Homo sapiens	TFII-I protein	730	79
1273	gi2440078	Homo sapiens	SPIN protein	730	79

266
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1274	AAB94010	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14130.	1921	91
1274	AAM79426	Homo sapiens	HYSE- Human protein SEQ ID NO 3072.	475	31
1274	AAM78442	Homo sapiens	HYSE- Human protein SEQ ID NO 1104.	455	32
1275	AAM93214	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 2616.	1353	99
1275	gi15292283	Drosophila melanogaster	LD45447p	172	30
1275	gi607003	Podospira anserina	beta transducin-like protein	87	23
1276	gi602438	Bos taurus	phosphoprotein	958	91
1276	AAM50706	Homo sapiens	DIAD- Human breast specific gene-encoded protein Sqmam042.	850	96
1276	gi7243755	Homo sapiens	neuronal phosphoprotein DARPP-32	508	100
1277	AAM47319	Homo sapiens	BODA- Human zinc finger structural domain 52.	2549	100
1277	gi15929040	Homo sapiens	Similar to RIKEN cDNA 3930402F13 gene	2549	100
1277	gi17391283	Homo sapiens	zinc finger protein 297	594	34
1278	AAB95788	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18745.	2742	99
1278	gi5579309	Mus musculus	zinc finger protein splice variant FIZ1-B	2308	86
1278	gi5579307	Mus musculus	zinc finger protein splice variant FIZ1-A	2308	86
1279	gi1016116	Cyanophora paradoxa	ribosomal protein L1	76	28
1279	gi 20834835 ref XP_136513.1	Mus musculus	similar to dJ1111A8.2 (novel protein with fibronectin type III domain)	79	29
1279	gi 11467315 ref NP_043172.1	Cyanophora paradoxa	ribosomal protein L1	76	28
1280	AAB94144	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14418.	811	99
1280	AAM42664	Homo sapiens	HUMA- Human kidney related polypeptide SEQ ID NO 533.	136	48
1280	AAM79272	Homo sapiens	HYSE- Human protein SEQ ID NO 1934.	136	48
1281	AAG74781	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5545.	533	99
1281	AAB53801	Homo sapiens	HUMA- Human colon cancer antigen protein sequence SEQ ID NO:1341.	307	95
1282	AAB93095	Homo sapiens	HELI- Human protein sequence SEQ ID NO:11945.	2618	99
1282	AAM93850	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3935.	1716	99
1282	AAU16242	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1195.	1677	100
1283	AAB94927	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16392.	806	100
1283	gi17979897	Homo sapiens	BAALC 1-5-6-7-8	74	31
1283	gi17979895	Homo sapiens	BAALC 1-5-6-8	74	31

267
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1284	gi1752638	Homo sapiens	RT14	630	100
1284	gi1703640	Homo sapiens	GCN5-like 1; GCN5L1	630	100
1284	gi2769585	Mus musculus	GCN5L1 protein	603	96
1285	AAAY77471	Homo sapiens	SCHE Human deubiquitinating protein Dub11, SEQ ID NO:34.	2455	87
1285	AAG64049	Homo sapiens	KAGA- Human deubiquitinating enzyme.	2448	88
1285	AAW30711	Homo sapiens	DAND Human ubiquitin-specific thiol protease DUB D38378.	2262	87
1286	gi13562000	Latrodectus geometricus	major ampullate spidroin 2	132	46
1286	gi545069	Mus sp.	Brn-3.2	128	42
1286	gi20249	Oryza sativa	gt-2	125	30
1287	ABB07304	Homo sapiens	SUGE- Human protein phosphatase SGP050.	2415	100
1287	AAB85358	Homo sapiens	INCY- Human phosphatase (PP) (clone ID 1723447CD1).	2161	98
1287	AAM39010	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2155.	929	45
1288	ABB07304	Homo sapiens	SUGE- Human protein phosphatase SGP050.	1333	99
1288	AAB85358	Homo sapiens	INCY- Human phosphatase (PP) (clone ID 1723447CD1).	1079	97
1288	AAM39010	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2155.	704	52
1289	gi10304393	Bombyx mori	ornithine decarboxylase antizyme	75	51
1289	gi17488611	Takifugu rubripes	Brain ankyrin 2	73	48
1289	gi8980338	Takifugu rubripes	FRANK2 protein	73	48
1290	gi12751397	Homo sapiens	EVG1	1106	100
1290	AAM95531	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4189.	219	100
1290	gi21064167	Drosophila melanogaster	AT27353p	212	29
1291	gi20988071	Mus musculus	Similar to RIKEN cDNA 2600011E07 gene	919	80
1291	gi13562004	Nephila madagascariensis	major ampullate spidroin 2-like protein	144	26
1291	gi2605967	Equine herpesvirus 4	24	136	21
1292	AAG75160	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5924.	361	73
1292	AAM95086	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 3744.	303	96
1292	gi13592175	Leishmania major	ppg3	270	24
1293	gi17932848	Homo sapiens	Spir-2 protein	1060	100
1293	gi20072945	Mus musculus	Similar to Spir-2 protein	932	85
1293	gi21449822	Mus musculus	Spir-2 protein	932	85
1294	AAU20195	Homo sapiens	HUMA- Human novel endocrine antigen, SEQ ID No 252.	587	91
1294	gi20151337	Drosophila melanogaster	GH04831p	481	46
1294	gi13548492	Caenorhabditis	ZC410.7b	349	43

268
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		elegans			
1295	gi181097	Homo sapiens	gamma-A-crystallin	968	99
1295	gi387135	Mus musculus	gamma-A-crystallin	845	84
1295	gi203627	Rattus norvegicus	gamma-A-crystallin	837	83
1296	AAG02474	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6555.	266	96
1296	gi7544050	Streptomyces coelicolor A3(2)	beta-N-acetylhexosaminidase	69	30
1296	gi10518509	Streptomyces plicatus	B-N-acetylhexosaminidase	69	33
1297	AAB92496	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10598.	966	38
1297	gi6581093	Mus musculus	transposase-like protein	910	36
1297	gi6581097	Homo sapiens	transposase-like protein	857	35
1298	gi12060820	Homo sapiens	serologically defined breast cancer antigen NY-BR-15	2775	100
1298	AAB95624	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18344.	963	35
1298	AAV07001	Homo sapiens	LUDW- Breast cancer associated antigen precursor sequence.	706	93
1299	AAU74387	Homo sapiens	CORI- Breast tumour-specific protein B305D fusion construct.	595	49
1299	AAU74379	Homo sapiens	CORI- Breast tumour-specific protein B11Ag1 isoform B11C-9,16.	595	49
1299	AAU74378	Homo sapiens	CORI- Breast tumour-specific protein B11Ag1 isoform B11C-9.	595	49
1300	AAG01406	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5487.	421	100
1300	gi21429922	Drosophila melanogaster	GH12382p	107	26
1300	gi5091552	Arabidopsis thaliana	T10024.21	101	29
1301	gi13111480	Homo sapiens	NeshBP	2268	89
1301	gi20071840	Mus musculus	Similar to DKFZP586L2024 protein	1851	81
1301	AAE01791	Homo sapiens	HUMA- Human gene 22 encoded secreted protein HOHDF66, SEQ ID NO:112.	1172	81
1302	AAM25877	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1392.	201	44
1302	AAE10329	Homo sapiens	INCY- Human transporter and ion channel-6 (TRICH-6) protein.	201	44
1302	AAZ11910_aa1	Homo sapiens	AXYS- Human potassium channel K+Hnov28 cDNA (5' splice variant 4).	183	34
1303	AAB28049	Homo sapiens	HUMA- Human secreted protein SEQ ID NO: 97.	204	100
1303	gi6715148	Mus musculus	selectively expressed in embryonic epithelia protein-1	72	31
1303	AAO10249	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 24141.	68	31
1304	AAV06060_aa1	Homo sapiens	UMIS Human imidazoline receptor subtype 1 cDNA.	91	32
1304	AAW43397	Homo sapiens	UMIS Human imidazoline receptor subtype 1 degradation product I2.	91	32
1304	AAW43396	Homo sapiens	UMIS Human imidazoline receptor	91	32

269
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			subtype 1.		
1305	gi19264036	Mus musculus	membrane-associated tyrosine-and threonine-specific cdc2-inhibitory kinase	94	26
1305	AAW56619	Homo sapiens	ONYX- Human cell growth regulatory protein.	90	26
1305	gi2460023	Homo sapiens	membrane-associated kinase	90	26
1306	gi1147611	Saccharomyces cerevisiae	Lpf3p	135	28
1306	ABB04721	Homo sapiens	SHAN- Human PP2464 protein SEQ ID NO:17.	101	22
1306	gi6630459	Arabidopsis thaliana	F23N19.12	93	39
1307	gi15277926	Homo sapiens	Similar to RIKEN cDNA 2700064H14 gene	1160	99
1307	gi20071138	Mus musculus	RIKEN cDNA 2700064H14 gene	1160	99
1307	AAM90766	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:18359.	719	71
1308	AAB93936	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13939.	813	95
1308	AAM00936	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 412.	313	52
1308	gi20502367	Homo sapiens	chaperone-ABC1-like	313	52
1309	ABB04628	Homo sapiens	BODA- Human lipoprotein precursor protein signal peptide 41 SEQ ID NO:2.	1719	100
1309	AAM95270	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 3928.	583	100
1309	AAW30650	Homo sapiens	GEMY Human secreted protein clone bg249 1 protein.	513	35
1310	AAE15854	Homo sapiens	ELIL Human SEZ6 mature protein.	75	24
1310	AAE15853	Homo sapiens	ELIL Human SEZ6 protein.	75	24
1310	AAU81976	Homo sapiens	INCY- Human secreted protein SECP2.	75	24
1311	gi560493	Rattus norvegicus	ribosomal protein L24	794	100
1311	gi3132823	Bos taurus	ribosomal protein L30	794	100
1311	gi292437	Homo sapiens	ribosomal protein L30	794	100
1312	gi21064771	Drosophila melanogaster	RH61467p	521	38
1312	gi18676554	Homo sapiens	FLJ00174 protein	391	34
1312	gi16182795	Drosophila melanogaster	GH12489p	365	33
1313	AAV84592	Homo sapiens	UNIW Amino acid sequennce of a human artemin polypeptide.	104	31
1313	gi17429323	Ralstonia solanacearum	PROBABLE GSPD-RELATED PROTEIN	95	29
1313	AAW87504	Homo sapiens	SIBI- Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA24.	93	31
1314	gi927637	Anthocidaris crassispina	dynein intermediate chain 2	604	29
1314	gi6594639	Homo sapiens	dynein intermediate chain DNAI1	602	28
1314	gi21040511	Homo sapiens	dynein, axonemal, intermediate	600	28

270

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			polypeptide, 1		
1315	gi20530222	Homo sapiens	BLOCK 25	1830	100
1315	gi11611545	Sus scrofa	aldose 1-epimerase	1670	89
1315	AAM40101	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3246.	1366	100
1316	AAM60715	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 32820.	2351	100
1316	gi5917666	Zea mays	extensin-like protein	148	24
1316	gi1749842	Yarrowia lipolytica	cell wall protein	120	35
1317	AAM25483	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:998.	975	99
1317	ABB11250	Homo sapiens	HYSE- Human Fc-gamma receptor homologue, SEQ ID NO:1620.	795	100
1317	AAG01757	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5838.	612	99
1318	gi15145797	Sus scrofa	basic proline-rich protein	532	29
1318	gi5917666	Zea mays	extensin-like protein	475	25
1318	gi5305335	Mycobacterium tuberculosis	proline-rich mucin homolog	428	25
1319	AAU11265	Homo sapiens	BODE- Human zinc finger protein 51.	1576	64
1319	gi15929737	Mus musculus	Similar to zinc finger protein 347	1557	56
1319	AAB93576	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12988.	1518	56
1320	AAM96103	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4761.	516	92
1320	gi17223819	Homo sapiens	transcription cofactor vestigial-like 2 protein	222	47
1320	AAM94333	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 2991.	158	35
1321	AAB95456	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17927.	1399	100
1321	gi21483586	Drosophila melanogaster	SD07343p	94	27
1321	gi18700713	Drosophila melanogaster	transcriptional corepressor Atro	94	27
1322	AAB94774	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15862.	3838	99
1322	gi18479137	Mus musculus	angiotensin	136	25
1322	gi2754696	Gallus gallus	high molecular mass nuclear antigen	134	21
1323	gi15929061	Homo sapiens	Similar to RIKEN cDNA 1500001K17 gene	1371	100
1323	gi18605654	Mus musculus	RIKEN cDNA 1500001K17 gene	1076	80
1323	gi9885296	Homo sapiens	LENG1 protein	1056	100
1324	gi292443	Homo sapiens	ribosomal protein S20	472	84
1324	gi57720	Rattus rattus	ribosomal protein S20 (AA 1-119)	472	84
1324	gi17932978	Homo sapiens	ribosomal protein S20	472	84
1325	AAB82946	Homo sapiens	UYNY Human androgen receptor trapped protein (ART).	661	30
1325	AAR23963	Homo sapiens	SNOW AFP-1 (Ala 2460 Val).	661	30
1325	AAR23962	Homo sapiens	SNOW AFP-1.	661	30
1326	gi12002024	Homo sapiens	brain my038 protein	371	100
1326	gi10178219	Arabidopsis thaliana	DNA-binding protein-like	97	41

271

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1326	gi19171280	Encephalitozoon cuniculi	similarity to HYPOTHETICAL ZINC FINGER PROTEIN YKF9_YEAST	92	31
1327	gi6566147	Drosophila melanogaster	large Forked protein	193	30
1327	AAU20496	Homo sapiens	HUMA- Human secreted protein, Seq ID No 488.	192	41
1327	AAU28174	Homo sapiens	HYSE- Novel human secretory protein, Seq ID No 343.	192	41
1328	AAB94745	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15792.	6782	99
1328	gi21392102	Drosophila melanogaster	RE22982p	867	27
1328	ABB89764	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2140.	624	97
1329	gi15991574	Homo sapiens	eferin	825	53
1329	gi17945944	Drosophila melanogaster	RE46851p	337	28
1329	gi2961133	Drosophila melanogaster	nuclear fallout	337	28
1330	AAM52332	Homo sapiens	BIOW- Human zinc finger protein 72.	3635	99
1330	gi14456629	Homo sapiens	dJ54B20.2 (novel KRAB box containing C2H2 type zinc finger protein)	3619	100
1330	gi15081398	Homo sapiens	kruppel-like zinc finger protein	1823	55
1331	gi13938317	Homo sapiens	Similar to zinc finger protein 202	416	43
1331	AAM40994	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5925.	393	43
1331	AAM39208	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2353.	393	43
1332	gi56952	Rattus norvegicus	put. preoptic regulatory factor-2	381	98
1332	gi2190355	Dictyostelium discoideum	racGAP	192	30
1332	gi6226761	Dictyostelium discoideum	class VII unconventional myosin	190	30
1333	gi21359654	Homo sapiens	LGI4	2398	93
1333	gi20975688	Homo sapiens	leucine-rich glioma inactivated protein 4	2398	93
1333	gi20975756	Mus musculus	leucine-rich glioma inactivated protein 4	2169	84
1334	gi15020655	Homo sapiens	ATP/GTP-binding protein	5570	99
1334	AAB94270	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14689.	4071	100
1334	gi21428364	Drosophila melanogaster	GM14375p	267	21
1335	gi11493544	Homo sapiens	PRO1460	412	100
1336	gi15079491	Homo sapiens	Similar to RIKEN cDNA 1200014H14 gene	1805	100
1336	gi13879544	Mus musculus	Similar to RIKEN cDNA 1200014H14 gene	933	88
1336	gi14043175	Homo sapiens	Similar to RIKEN cDNA 1200014H14 gene	920	100
1337	gi20988758	Mus musculus	RIKEN cDNA 1700030A21 gene	679	84
1337	AAB67449	Homo sapiens	INCY- Amino acid sequence of a human chaperone polypeptide.	566	100

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1337	gi9757767	Arabidopsis thaliana	tetratricopeptide repeat protein 2-like	163	39
1338	gi18265515	Mus musculus	LIX1	1364	94
1338	gi18265513	Gallus gallus	LIX1	1203	83
1338	AAM93000	Homo sapiens	HUMA- Human digestive system antigen SEQ ID NO: 2349.	753	87
1339	gi14024367	Mesorhizobium loti	weak similarity to ubiquinone/menaquinone biosynthesis methyl transferase	185	32
1339	gi20804035	Mesorhizobium loti	HYPOTHETICAL PROTEIN	148	29
1339	gi21107889	Xanthomonas axonopodis pv. citri str. 306	methyltransferase	115	33
1340	gi559237	Petroselinum crispum	tyrosine-rich hydroxyproline-rich glycoprotein	90	26
1340	gi4406393	Bos taurus	differentiation enhancing factor 1	88	35
1340	gi17939907	Pseudorabies virus	UL36 protein	84	32
1341	ABB11819	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:2189.	2334	97
1341	AAW80398	Homo sapiens	GEMY A secreted protein encoded by clone cw1543_3.	1887	98
1341	gi7547029	Homo sapiens	GAP-like protein	347	33
1342	AAB50867	Homo sapiens	ZYMO Human zalpha31.	701	100
1342	AAB54229	Homo sapiens	HUMA- Human pancreatic cancer antigen protein sequence SEQ ID NO:681.	701	100
1342	AAB10232	Homo sapiens	GEMY Human fetal placenta protein fragment AP162_1i.	361	97
1344	AAV73346	Homo sapiens	INCY- HTRM clone 619699 protein sequence.	1335	60
1344	AAB43912	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:1357.	1334	60
1344	gi498721	Homo sapiens	zinc finger protein	1047	51
1345	gi13516895	Homo sapiens	MLZE	2166	100
1345	gi13516897	Mus musculus	MLZE	875	47
1345	AAB93904	Homo sapiens	HELI- Human protein sequence SEQ ID NO:13862.	360	30
1346	ABB12463	Homo sapiens	HYSE- Human bone marrow expressed protein SEQ ID NO: 302.	556	100
1346	AAM77500	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 37806.	342	100
1346	AAM64730	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 36835.	342	100
1347	gi12751092	Homo sapiens	PNAS-123	306	98
1347	gi9988480	Ursus maritimus	luteinizing hormone receptor	72	31
1347	gi1196882	Ovis aries	luteinizing hormone receptor	69	31
1348	gi12018147	Chlamydomonas reinhardtii	vegetative cell wall protein gp1	227	27
1348	gi15145793	Sus scrofa	basic proline-rich protein	221	28
1348	gi6523547	Volvox carteri f. nagariensis	hydroxyproline-rich glycoprotein DZ-HRGP	220	30

273
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1349	AAB95039	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16797.	2187	100
1349	gi9963804	Homo sapiens	zinc finger protein ZNF286	1038	46
1349	AAM39130	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2275.	1015	46
1350	AAM49164	Homo sapiens	BODE- Human cytochrome bcl compound core protein II10.23.	471	100
1350	gi1905852	Drosophila melanogaster	reverse transcriptase polyprotein	98	25
1350	gi2429118	Leishmania major	DDI1; L3169.2	95	27
1351	gi19354550	Mus musculus	similar to src homology three (SH3) and cysteine rich domain	1897	92
1351	gi1799566	Mus musculus	stac	788	47
1351	AAW59642	Homo sapiens	SHIO Amino acid sequence of human Stac protein.	775	45
1352	AAU17301	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 866.	1347	100
1352	AAE11776	Homo sapiens	INCY- Human kinase (PKIN)-10 protein.	1204	100
1352	AAU17300	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 865.	618	100
1353	gi3986121	Rattus norvegicus	synaptopodin	78	28
1353	gi153373	Streptomyces griseus	ORFC	78	31
1353	gi5002552	Streptomyces griseus subsp. griseus	NonC	78	31
1354	gi19353360	Mus musculus	RIKEN cDNA 2010000G05 gene	260	56
1354	AAO13877	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 27769.	259	54
1354	AAB58207	Homo sapiens	ROSE/ Lung cancer associated polypeptide sequence SEQ ID 545.	259	54
1355	ABB08788	Homo sapiens	BODE- Human LAMC3 69.85 SEQ ID NO 2.	588	32
1355	AAM79102	Homo sapiens	HYSE- Human protein SEQ ID NO 1764.	588	32
1355	gi20071167	Mus musculus	Similar to Per1 interacting protein	558	30
1356	gi1839334	Homo sapiens	ribosomal protein L37	510	96
1356	gi57121	Rattus norvegicus	ribosomal protein L37	510	96
1356	gi461232	Homo sapiens	ribosomal protein L37	510	96
1357	gi20988686	Homo sapiens	DMRT-like family C1	553	100
1357	AAM42520	Homo sapiens	HUMA- Human kidney related polypeptide SEQ ID NO 389.	386	89
1357	AAM99705	Homo sapiens	HUMA- Human excretory related polypeptide SEQ ID NO 442.	386	89
1358	gi339783	Homo sapiens	slow skeletal muscle troponin T	1205	100
1358	gi546023	Homo sapiens	troponin T slow isoform; TnT slow isoform	1205	100
1358	gi4056562	Homo sapiens	slow skeletal muscle troponin T	1205	100
1359	AAU16178	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1131.	2603	98
1359	gi21464470	Drosophila	RH63657p	487	37

274
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		melanogaster			
1359	AAM65817	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26123.	389	100
1360	AAG89164	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 284.	1774	99
1360	ABA05334_aa1	Homo sapiens	MILL- Human fucosyltransferase family member 32132 coding sequence.	853	42
1360	AAM47905	Homo sapiens	MILL- Human fucosyltransferase family member 32132.	853	42
1361	gi14529886	Mus musculus	bM145O4.1 (novel protein)	1176	47
1361	AAU15865	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 818.	1091	46
1361	AAE02779	Homo sapiens	CURA- Human PRO-C-MG.45 protein encoded by DNA-C-MG.45-1776 cDNA clone.	1080	46
1362	gi2323287	multiple sclerosis associated retrovirus	polyprotein	3104	81
1362	gi 331995 gb AAB03091.1	AKV murine leukemia virus	gag-pol polyprotein (tag amber codon at 2250-2252 inserts Gln in Mo-MuLV)	1368	37
1362	gi 535518 gb AAA92679.1	Murine leukemia virus	gag-pol polyprotein	1366	36
1363	AAB92554	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10741.	1062	100
1363	AAM91124	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:18717.	466	97
1363	gi1141790	Drosophila melanogaster	nonmuscle myosin-II heavy chain	337	25
1364	gi19913468	Homo sapiens	similar to zinc finger protein	519	98
1364	AAM78974	Homo sapiens	HYSE- Human protein SEQ ID NO 1636.	321	64
1364	gi5679576	Homo sapiens	zinc finger 41	321	64
1365	gi15488932	Homo sapiens	Similar to RIKEN cDNA 1700037B15 gene	1002	100
1365	gi17385767	Mus musculus	RTP801-like protein	935	93
1365	gi17432247	Mus musculus	SMHS1	933	93
1366	AAM39808	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2953.	105	24
1366	AAB47142	Homo sapiens	INCY- CDIFF-24, Incyte ID No. 1546633CD1.	101	24
1366	AAM41594	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6525.	98	24
1367	gi6409379	Homo sapiens	zinc finger protein ZNF229	2299	100
1367	gi10864174	amino acids 1-420] [Homo sapiens	ZNF229	2299	100
1367	AAB95278	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17486.	1973	48
1368	AAG74304	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5068.	933	100

275
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1368	AAG04049	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8130.	487	90
1368	gi161292	Loligo pealei	neurofilament protein	96	22
1369	gi21314907	Homo sapiens	PYRIN-containing APAF1-like Protein 7	4262	89
1369	AAE07514	Homo sapiens	MILL- Human PYRIN-1 protein.	2156	49
1369	gi18182339	Homo sapiens	PYRIN-containing APAF1-like protein 1	2156	49
1370	gi13183793	Homo sapiens	CECR2 protein	6203	97
1370	AAM70233	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 30539.	1572	100
1370	AAM67858	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28164.	1572	100
1371	gi10800858	Homo sapiens	aminopeptidase B	3476	99
1371	gi15082509	Homo sapiens	arginyl aminopeptidase (aminopeptidase B)	3437	98
1371	gi10933784	Homo sapiens	aminopeptidase B	3433	98
1372	AAE15250	Homo sapiens	INCY- Human RNA metabolism protein-13 (RMEP-13).	2339	99
1372	gi20072721	Mus musculus	RIKEN cDNA 3100004P22 gene	2223	91
1372	gi2098575	Homo sapiens	F25451_2	1602	99
1373	gi6841246	Homo sapiens	HSPC298	528	90
1373	gi2384748	peach rosette mosaic virus	RNA1 polyprotein	71	33
1373	gi 6841246 gb AAF28976.1 AF161416_1	Homo sapiens	HSPC298	528	90
1374	gi532113	Caenorhabditis elegans	homeotic region most like HMPB_DROME: homeotic proboscipedia protein	211	21
1374	gi18447198	Drosophila melanogaster	GH09355p	209	19
1374	gi6984160	Streptococcus cristatus	srpA	190	17
1375	gi21430538	Drosophila melanogaster	LP11564p	662	53
1375	AAM41000	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5931.	282	27
1375	AAM40999	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 5930.	282	27
1376	AAG66311	Homo sapiens	BIOD- Human zinc finger protein 46.	2311	99
1376	gi19548796	Mus musculus	multifinger protein mKr2	2209	85
1376	gi4567180	Homo sapiens	BC37295_2 (partial)	1765	64
1377	AAU16960	Homo sapiens	HUMA- Human novel secreted protein, SEQ ID 201.	3079	98
1377	gi19263612	Mus musculus	Similar to expressed sequence AI413166	1862	87
1377	AAM67519	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 27825.	808	100
1378	ABB05653	Homo sapiens	BODE- Human DNA binding protein RFX2-89 SEQ ID NO:2.	3952	93

276

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1378	AAB94287	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14728.	3143	99
1378	gi11994733	Arabidopsis thaliana	contains similarity to cell wall-plasma membrane linker protein~gene id:MKA23.5	347	25
1379	AAI66827_aa1	Homo sapiens	MILL- Human protein kinase polypeptide 23546 coding sequence.	3993	100
1379	AAG65766	Homo sapiens	MILL- Human protein kinase polypeptide 23546.	3993	100
1379	gi18734	Glycine max	DNA-directed RNA polymerase	147	30
1380	gi847724	Homo sapiens	methylthioadenosine phosphorylase	1464	95
1380	gi13111476	Mus musculus	methylthioadenosine phosphorylase	1377	90
1380	gi13277990	Mus musculus	RIKEN cDNA 1300019I21 gene	1377	90
1381	gi12584841	Homo sapiens	HT036	831	94
1381	gi12584839	Homo sapiens	HT036-ISO	831	94
1381	AAG03594	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7675.	403	88
1382	gi12584839	Homo sapiens	HT036-ISO	986	85
1382	gi12584841	Homo sapiens	HT036	809	83
1382	AAG03594	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7675.	381	70
1383	gi12584839	Homo sapiens	HT036-ISO	916	99
1383	gi12584841	Homo sapiens	HT036	739	99
1383	gi17427028	Ralstonia solanacearum	CONSERVED HYPOTHETICAL PROTEIN	367	46
1384	AAM00883	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 359.	574	99
1384	AAM00770	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 133.	573	99
1384	gi17862960	Drosophila melanogaster	SD05477p	349	54
1385	AAB95751	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18660.	173	32
1385	AAV16782	Homo sapiens	GEMY Human secreted protein (clone cg426_8).	142	50
1385	AAM86749	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:14342.	118	32
1386	gi19879436	Homo sapiens	COASTER	5431	95
1386	AAB92536	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10701.	2682	100
1386	AAU16460	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1413.	1864	99
1387	gi6103649	Homo sapiens	F-box protein FBX10	2878	99
1387	AAV83080	Homo sapiens	UYNY F-box protein FBP-12.	963	100
1387	gi6164739	Homo sapiens	F-box protein Fbx10	963	100
1388	gi15082426	Homo sapiens	Similar to RIKEN cDNA 2810055F11 gene	1840	99
1388	gi13435795	Mus musculus	Similar to RIKEN cDNA 2810055F11 gene	1661	88
1388	AAU27647	Homo sapiens	ZYMO Human protein AFP301973.	673	99
1389	AAM38689	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 1834.	2085	66
1389	gi38032	Homo sapiens	ZNF43	2085	66
1389	gi16306806	Homo sapiens	zinc finger protein 43 (HTF6)	2085	66

277
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1390	AAB95279	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17488.	3194	100
1390	gi14329547	Homo sapiens	dJ1104A8.1 (novel protein)	629	99
1390	gi8250181	Drosophila melanogaster	D-Titin	160	20
1391	AAM50137	Homo sapiens	CYTO- Human kinesin motor protein HsKip3b motor domain.	972	85
1391	gi12862607	Mus musculus	kinesin superfamily protein 19A	663	94
1391	ABB07410	Homo sapiens	CYTO- Human kinesin motor protein, HsKip3A.	646	48
1392	gi19484159	Mus musculus	Similar to RIKEN cDNA 5430428K15 gene	807	79
1392	AAB94842	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16016.	401	98
1392	gi3875383	Caenorhabditis elegans	D2085.2	121	29
1393	gi13442786	Mus musculus	Drctnnb1a	773	70
1393	gi17511709	Homo sapiens	down-regulated by Ctnnb1, a	765	66
1393	gi13442784	Homo sapiens	DRCTNNB1A	752	66
1394	ABB08360	Homo sapiens	GENZ Human eIF3 amino acid sequence.	1785	95
1394	gi2351380	Homo sapiens	translation initiation factor eIF3 p40 subunit	1785	95
1394	gi12653235	Homo sapiens	eukaryotic translation initiation factor 3, subunit 3 (gamma, 40kD)	1785	95
1395	gi3169261	Mus musculus	T-box transcription factor	1584	98
1395	gi3128382	Mus musculus	MmTbx14	1398	98
1395	gi21449340	Danio rerio	T-box containing transcription factor	1366	85
1396	AAB95361	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17667.	702	99
1396	gi19071496	Pseudomonas syringae pv. maculicola	type III effector HopPmaI	100	31
1396	gi7140837	Streptomyces seoulensis	dihydrolipoamide acetyltransferase	98	34
1397	gi12655165	Homo sapiens	zinc finger protein 256	2656	100
1397	AAZ35253_aal	Homo sapiens	UYSH- Human leucine zipper protein gene (BMH) (CBLAYD06).	2646	99
1397	AAZ32427	Homo sapiens	UYSH- Human leucine zipper protein CBLAYD06.	2646	99
1398	gi458692	Homo sapiens	homologous to mouse gene PC326:GenBank Accession Number M95564	2043	67
1398	gi200241	Mus musculus	protein PC326	1600	59
1398	AAM73935	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 34241.	818	100
1399	gi17223819	Homo sapiens	transcription cofactor vestigial-like 2 protein	697	95
1399	AAM94333	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 2991.	156	54
1399	gi1809231	Homo sapiens	coded for by human cDNAs R76043 (NID:g850725), R65857 (NID:g838495) and H12868 (NID:g877688)	156	54

278
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1400	gi18874272	Homo sapiens	axonemal dynein heavy chain DNAH5	5798	97
1400	gi18449111	Mus musculus	axonemal dynein heavy chain 5	5549	92
1400	gi14335444	Homo sapiens	axonemal dynein heavy chain 8	4341	70
1401	AAG03810	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7891.	760	92
1401	gi186800	Homo sapiens	ribosomal protein L12	760	92
1401	gi17390751	Mus musculus	ribosomal protein L12	759	91
1402	gi17512582	Homo sapiens	Similar to RIKEN cDNA 2700059L22 gene	1611	100
1402	AAM95231	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 3889.	698	97
1402	gi18447648	Drosophila melanogaster	SD06318p	161	22
1403	AAU20514	Homo sapiens	HUMA- Human secreted protein, Seq ID No 506.	255	52
1403	AAV54057	Homo sapiens	GEHO Amino acid sequence of novo DNA cytosine methyltransferase DNMT3A.	143	19
1403	gi295671	Saccharomyces cerevisiae	selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependant RNA polymerase I and III	143	19
1404	gi13938351	Homo sapiens	Similar to zinc finger protein 268	3615	99
1404	gi186774	Homo sapiens	zinc finger protein	2626	55
1404	gi2739353	Homo sapiens	ZNF91L	2579	54
1405	AAM95368	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4026.	792	100
1405	AAV30795	Homo sapiens	USSH A human trichohyalin (TRHY) protein.	128	22
1405	gi292836	Homo sapiens	trichohyalin	128	22
1406	AAW41936	Homo sapiens	GEMY Secreted protein AS152_1.	177	100
1406	AAW42082	Homo sapiens	GEMY The amino acid sequence of the AS152_1 protein.	177	100
1406	gi19908760	Aspergillus nomius	AFLR	79	31
1407	AAB94541	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15286.	1020	100
1407	gi21324755	Corynebacterium glutamicum ATCC 13032	Translation initiation factor 2 (GTPase)	124	30
1407	gi12044832	Chlamydomonas reinhardtii	DEAH-box RNA helicase	124	33
1408	gi13469801	Homo sapiens	tandem FYVE fingers-1 protein	1874	100
1408	gi10834636	Homo sapiens	double FYVE-containing protein 1	1874	100
1408	gi18369779	Homo sapiens	phosphoinositide-binding protein	1874	100
1409	ABB04291	Homo sapiens	UYNA- Human testis-specific protein.	2586	100
1409	gi12330993	Homo sapiens	testis-specific protein TSP-NY	2586	100
1409	AAM95389	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4047.	1477	98
1410	gi19168684	Encephalitozoon cuniculi	similarity to HYPOTHETICAL PROTEIN YANE_SCHPO	262	33
1410	gi476061	Saccharomyces cerevisiae	YBR0834	237	29
1410	gi3237301	Saccharomyces cerevisiae	Vid24p	237	29
1411	AAM76193	Homo sapiens	MOLE- Human bone marrow	238	100

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			expressed probe encoded protein SEQ ID NO: 36499.		
1411	AAM63381	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 35486.	238	100
1411	gi5669894	Human herpesvirus 8	latent nuclear antigen	216	23
1412	gi19484159	Mus musculus	Similar to RIKEN cDNA 5430428K15 gene	955	80
1412	AAB94842	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16016.	406	100
1412	gi3875383	Caenorhabditis elegans	D2085.2	167	29
1413	gi15080358	Homo sapiens	Similar to LOC88745	1289	98
1413	gi6841156	Homo sapiens	HSPC253	912	99
1413	gi16769400	Drosophila melanogaster	LD29276p	320	34
1414	AAV28589	Homo sapiens	DAND TIA-1 binding protein 2 (TIABP2).	2860	95
1414	AAW64717	Homo sapiens	DAND Human TIABP2 protein.	2860	95
1414	AAR72830	Homo sapiens	DAND Human TIABP2.	2860	95
1415	AAH23820 aa1	Homo sapiens	INCY- Human transferase HTFS-20 cDNA, SEQ ID NO:62.	1252	100
1415	AAU11627	Homo sapiens	BODA- Human zinc finger protein 26.	1252	100
1415	AAB73513	Homo sapiens	INCY- Human transferase HTFS-20, SEQ ID NO:20.	1252	100
1416	gi1163174	Rattus norvegicus	similar to yeast Sec6p, Swiss-Prot Accession Number P32844; similar to mammalian B94, Swiss-Prot Accession Number Q03169; Method: conceptual translation supplied by author	386	25
1416	gi179331	Homo sapiens	B94 protein	373	26
1416	AAV51115	Homo sapiens	INCY- Human HSEC6 protein.	366	24
1417	gi10440416	Homo sapiens	FLJ00043 protein	6728	95
1417	gi13604143	Mus musculus	tangerin A	4044	57
1417	AAM00864	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 340.	1947	98
1418	AAU17059	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 624.	2491	99
1418	gi18026290	Homo sapiens	Gab3	2491	99
1418	gi18026288	Mus musculus	Gab3	1901	76
1419	AAU17059	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 624.	2886	99
1419	gi18026290	Homo sapiens	Gab3	2886	99
1419	gi18026288	Mus musculus	Gab3	2178	74
1420	gi13528684	Homo sapiens	Similar to ribosomal protein S6 kinase, 52kD, polypeptide 1	2737	100
1420	AAB65612	Homo sapiens	SUGE- Novel protein kinase, SEQ ID NO: 138.	2695	94
1420	AAM41924	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6855.	546	55
1421	gi2330003	Gallus gallus	glutamine rich protein	151	26
1421	gi15292043	Drosophila melanogaster	LD36157p	141	23
1421	gi159333	Leishmania	glycoprotein 96-92	132	25

280
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		major			
1422	AAE06617	Homo sapiens	SAGA Human protein having hydrophobic domain, HP10750.	970	89
1422	AAM25467	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:982.	970	89
1422	gi15292607	Drosophila melanogaster	SD10847p	182	25
1423	AAM93462	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3124.	1284	100
1423	gi3219288	Mus musculus	PA28 gamma subunit	1284	100
1423	gi2351200	Mus musculus	Ki antigen	1284	100
1424	gi21104462	Homo sapiens	OK/SW-CL.30	1926	100
1424	gi20805945	Drosophila melanogaster	wingful	675	45
1424	gi20269077	Drosophila melanogaster	Notum protein	675	45
1425	gi15721860	Ciona intestinalis	dynein intermediate chain 3	739	49
1425	gi1817526	Anthocidaris crassispina	intermediate chain 1	735	46
1425	gi7580490	Homo sapiens	NM23-H8	473	40
1426	AAU81984	Homo sapiens	INCY- Human secreted protein SECP10.	454	77
1426	gi6580722	Yersinia pseudotuberculosis	O-unit polymerase-like protein	82	25
1426	gi6735326	Arabidopsis thaliana	phenylalanine-tRNA synthetase-like protein	76	30
1427	ABB11972	Homo sapiens	HYSE- Human VM106R.1 homologue, SEQ ID NO:2342.	698	60
1427	gi16648066	Drosophila melanogaster	GH08630p	500	39
1427	gi3880445	Caenorhabditis elegans	contains similarity to Pfam domain: PF02214 (K ⁺ channel tetramerisation domain), Score=79.5, E-value=2.3e-20, N=1	317	37
1428	AAE00674	Homo sapiens	HUMA- Human protein tyrosine kinase receptor (PTK) from clone HSSJQ45.	1302	100
1428	AAM39039	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2184.	1107	55
1428	gi16648322	Drosophila melanogaster	LD29155p	514	37
1429	gi12803785	Homo sapiens	Similar to cell cycle progression 2 protein	2341	100
1429	AAB94081	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14279.	2334	99
1429	gi2352902	Homo sapiens	cell cycle progression 2 protein	1686	94
1430	gi21040522	Homo sapiens	similar to kinase D-interacting substance of 220 kDa; ankyrin repeat-rich membrane-spanning protein	1511	99
1430	AAU20586	Homo sapiens	HUMA- Human secreted protein, Seq ID No 578.	1387	100
1430	AAE01031	Homo sapiens	HUMA- Human death domain-containing receptor (DDCR) protein from HFIHQ20 clone.	1259	100

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1431	gi11875764	Homo sapiens	neurogenin 2	1301	100
1431	gi1504095	Mus musculus	DNA-binding protein	1146	83
1431	gi1666910	Mus musculus	neurogenin 2	1138	83
1432	gi19067867	Pseudonaja textilis	pseudonajatoxin b precursor	74	34
1432	gi5230714	Pseudonaja textilis	long neurotoxin precursor	74	34
1432	gi14250573 gb AAH08741.1 AAH08741	Homo sapiens	LIM protein (similar to rat protein kinase C-binding enigma)	67	40
1433	gi1881738	Acanthamoeba castellanii	myosin-I binding protein Acan125	366	28
1433	gi14701866	Dictyostelium discoideum	carmil	325	25
1433	gi20087129	Dictyostelium discoideum	DEVELOPMENTAL PROTEIN DG1112	139	23
1434	gi12584159	Homo sapiens	zinc finger protein 268	5261	100
1434	gi14579579	Homo sapiens	ZNF268B	4395	100
1434	gi19851928	Homo sapiens	CLL-associated antigen KW-4 splice variant 2	4342	98
1435	ABN00011_aal	Homo sapiens	AEOM- Human genome-derived myosin-like protein encoding cDNA SEQ ID NO:2.	2623	98
1435	ABB06333	Homo sapiens	AEOM- Human genome-derived myosin-like protein SEQ ID NO:3.	2623	98
1435	gi15718364	Homo sapiens	myosin heavy chain	2623	98
1436	gi17945712	Drosophila melanogaster	RE31178p	360	40
1436	gi18491231	Arabidopsis thaliana	At2g41750/T11A7.15	167	26
1436	gi987652	Escherichia coli	hypothetical protein in UNG 3' region	116	25
1437	AAM79289	Homo sapiens	HYSE- Human protein SEQ ID NO 1951.	470	84
1437	AAG03927	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8008.	470	84
1437	AAB57110	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1688.	470	84
1438	AAW29658	Homo sapiens	GEMY Homo sapiens BH272_3 clone secreted protein.	829	89
1438	gi2282038	Homo sapiens	p21-Arc	829	89
1438	gi6983853	Mus musculus	Arp2/3 complex subunit p21-Arc	814	87
1439	ABB50184	Homo sapiens	INCY- Human transcription factor TRFX-35.	1459	57
1439	AAB93164	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12091.	1425	65
1439	AAM40308	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3453.	1418	51
1440	AAU27656	Homo sapiens	ZYMO Human protein AFP548753.	537	99
1440	AAU21102	Homo sapiens	HUMA- Human novel foetal antigen, SEQ ID NO 1346.	337	100
1440	AAM75756	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36062.	313	100
1441	gi9294050	Arabidopsis	protein kinase-like protein	106	30

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		thaliana			
1441	gi15983765	Arabidopsis thaliana	AT3g24550/MOB24_8	106	30
1441	gi13877617	Arabidopsis thaliana	protein kinase-like protein	106	30
1442	gi11064131	Homo sapiens	bB329D4.1.2 (novel protein, isoform 2)	403	97
1442	gi11064132	Homo sapiens	bB329D4.1.1 (novel protein, isoform 1)	383	89
1442	gi11064130	Homo sapiens	bB329D4.1.3 (novel protein, isoform 3)	374	98
1443	gi19347650	Homo sapiens	BNIP-Salpa	1386	100
1443	gi20380760	Homo sapiens	BNIP-2 similar	1386	100
1443	gi19347652	Homo sapiens	BNIP-Sbeta	1082	99
1444	gi2429362	Santalum album	proline rich protein	144	31
1444	gi167728	Dictyostelium discoideum	trans-acting factor	127	34
1444	gi4539386	Arabidopsis thaliana	extensin-like protein	122	27
1445	gi17223626	Homo sapiens	ATP-binding cassette A10	378	100
1445	AAU18378	Homo sapiens	HUMA- Human endocrine polypeptide SEQ ID No 333.	277	72
1445	gi17223624	Homo sapiens	ATP-binding cassette A9	277	72
1446	gi633990	Mus musculus	zinc finger protein	1964	86
1446	gi4481920	Mus musculus	Ozrf1 protein	1961	86
1446	gi1336158	Rattus norvegicus	pancreas only zinc finger protein	1954	85
1447	gi13186114	Homo sapiens	rab interacting lysosomal protein	1850	92
1447	gi4539560	Streptomyces coelicolor A3(2)	integral membrane protein with kinase activity	149	26
1447	gi16648230	Drosophila melanogaster	GH23825p	139	22
1448	gi13186114	Homo sapiens	rab interacting lysosomal protein	2047	99
1448	gi4539560	Streptomyces coelicolor A3(2)	integral membrane protein with kinase activity	149	26
1448	gi9858781	Lycopersicon esculentum	BAC19.13	139	23
1449	AAE01693	Homo sapiens	HUMA- Human gene 22 encoded secreted protein HOFNM53, SEQ ID NO:105.	314	67
1449	ABB90444	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2820.	115	53
1449	AAB50967	Homo sapiens	GETH Human PRO1410 protein.	115	53
1450	gi12002226	Homo sapiens	C3HC4-type zinc finger protein	2042	99
1450	gi16751522	Mus musculus	dioxin inducible factor 3	1966	88
1450	gi13752272	Homo sapiens	laryngeal carcinoma related protein 1	1504	99
1451	gi13752754	Homo sapiens	zinc finger 1111	1977	59
1451	gi10440398	Homo sapiens	FLJ00032 protein	1893	58
1451	AAB95862	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18929.	1881	60
1452	gi11064132	Homo sapiens	bB329D4.1.1 (novel protein, isoform 1)	380	94
1452	gi11064131	Homo sapiens	bB329D4.1.2 (novel protein, isoform 2)	370	97
1452	gi11064130	Homo sapiens	bB329D4.1.3 (novel protein, isoform 3)	369	98

283
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			3)		
1453	AAB63251	Homo sapiens	LUDW- Human breast cancer associated antigen protein sequence SEQ ID NO:613.	964	99
1453	AAB63250	Homo sapiens	LUDW- Human breast cancer associated antigen protein sequence SEQ ID NO:612.	415	74
1453	AAB63249	Homo sapiens	LUDW- Human breast cancer associated antigen protein sequence SEQ ID NO:611.	380	74
1454	AAM80207	Homo sapiens	HYSE- Human protein SEQ ID NO 3853.	2233	56
1454	AAM79223	Homo sapiens	HYSE- Human protein SEQ ID NO 1885.	2233	56
1454	gi5080758	Homo sapiens	BC331191_1	2159	52
1455	gi189044	Homo sapiens	zinc finger protein 42	368	30
1455	gi8886436	Homo sapiens	myeloid zinc finger protein 1 splice variant	368	30
1455	gi4567181	Homo sapiens	BC37295_3	332	26
1456	gi17984559	Brucella melitensis	PUTATIVE DNA-BINDING PROTEIN	184	36
1456	gi15159614	Agrobacterium tumefaciens str. C58 (Cereon)	AGR_L_2275p	180	37
1456	gi15076209	Sinorhizobium meliloti	HYPOTHETICAL PROTEIN	176	34
1457	AAB21018	Homo sapiens	INCY- Human nucleic acid-binding protein, NuABP-22.	1085	50
1457	gi55471	Mus musculus	Zfp-29	713	49
1457	AAU16006	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 959.	688	50
1458	gi16877762	Homo sapiens	Similar to cAMP responsive element modulator	672	100
1458	gi1110452	Homo sapiens	hCREM 1alpha protein	594	99
1458	gi30217	Homo sapiens	cAMP responsive element modulator	589	98
1459	AAB54325	Homo sapiens	HUMA- Human pancreatic cancer antigen protein sequence SEQ ID NO:777.	751	98
1459	gi18676566	Homo sapiens	FLJ00180 protein	231	28
1459	gi20467213	Homo sapiens	CARD15-like protein	219	32
1460	AAB94838	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16008.	2706	100
1460	AAB67573	Homo sapiens	INCY- Amino acid sequence of a human hydrolytic enzyme HYENZ5.	2701	99
1460	gi18605651	Mus musculus	Similar to RIKEN cDNA 4930584N22 gene	2220	82
1461	AAG03605	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7686.	299	98
1462	AAM00936	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 412.	1385	58
1462	gi20502367	Homo sapiens	chaperone-ABC1-like	1385	58
1462	AAM93752	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3735.	1381	57
1463	gi1212953	Homo sapiens	3-methyl-adenine DNA glycosylase	1470	99
1463	gi20162529	Homo sapiens	N-methylpurine-DNA glycosylase	1470	99

284
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1463	gi14336679	Homo sapiens	N-methylpurine-DNA	1470	99
1464	AAB56919	Homo sapiens	ROSE/ Human prostate cancer antigen protein sequence SEQ ID NO:1497.	1651	99
1464	gi8572544	Homo sapiens	3-methyl-adenine DNA glycosylase	1532	99
1464	gi20162529	Homo sapiens	N-methylpurine-DNA glycosylase	1505	100
1465	gi1912072	Mus musculus	AZ-1	3875	73
1465	gi1199931	Mus musculus	pre-acrosome localization protein	3875	73
1465	gi15079588	Homo sapiens	Similar to 5-azacytidine induced gene 1	3681	99
1466	AAU11813	Homo sapiens	UYLE- Cancer and neurogenesis associated gene, variant 5G-3V1.	3219	100
1466	AAU11817	Homo sapiens	UYLE- Cancer and neurogenesis associated gene, variant 5R23V2.	3144	98
1466	AAU11816	Homo sapiens	UYLE- Cancer and neurogenesis associated gene, variant 5R-3V2.	3144	98
1467	AAU12206	Homo sapiens	GETH Human PRO4979 polypeptide sequence.	1409	99
1467	AAG02178	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6259.	270	45
1467	gi470322	Daucus carota	proline-rich protein	99	30
1468	gi18676488	Homo sapiens	FLJ00141 protein	605	96
1468	AAG01457	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5538.	372	100
1468	gi15209365	Xenopus laevis	kinesin-like protein	164	34
1469	gi18676608	Homo sapiens	FLJ00203 protein	1829	100
1469	gi20149221	Homo sapiens	axonemal heavy chain dynein type 3	1550	41
1469	gi17225486	Homo sapiens	ciliary dynein heavy chain 7	1521	41
1470	gi17425160	Mus musculus	testis serine protease2	373	60
1470	gi15320416	Mus musculus	serine protease-like 1	340	61
1470	gi21064337	Drosophila melanogaster	RE07247p	260	48
1471	AAE19166	Homo sapiens	INCY- Human protease, PRTS-3 protein.	1327	99
1471	gi9957760	Homo sapiens	KLK15	1327	99
1471	gi14484922	Homo sapiens	prostinogen	1323	99
1472	AAB94615	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15462.	2500	99
1472	AAAY51846	Homo sapiens	PFEF/ Human 18.1 homolog protein fragment.	1382	100
1472	AAM40314	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3459.	572	26
1473	gi18606245	Mus musculus	RIKEN cDNA 1810020E01 gene	712	70
1473	AAM38641	Homo sapiens	HUMA- Human colorectal cancer antigen SEQ ID NO: 156.	692	96
1473	AAM93170	Homo sapiens	HUMA- Human digestive system antigen SEQ ID NO: 2519.	692	96
1474	AAU82017	Homo sapiens	INCY- Human secreted protein SECP43.	4623	97
1474	gi18676716	Homo sapiens	FLJ00257 protein	4041	99
1474	AAM00753	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 116.	2042	72
1475	AAU82017	Homo sapiens	INCY- Human secreted protein SECP43.	4372	93
1475	gi18676716	Homo sapiens	FLJ00257 protein	3769	94
1475	AAB47134	Homo sapiens	INCY- CDIFF-15, Incyte ID No.	1757	48

285
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			3478571CD1.		
1476	ABB90752	Homo sapiens	UYJO Human Tumour Endothelial Marker polypeptide SEQ ID NO 236.	1138	78
1476	gi6708478	Mus musculus	formin-like protein	784	57
1476	gi4101720	Mus musculus	lymphocyte specific formin related protein	744	60
1477	gi18848216	Homo sapiens	Similar to RIKEN cDNA 2410016C14 gene	732	97
1477	AAG02288	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6369.	561	100
1477	gi4234	Saccharomyces cerevisiae	the product of PRP22 gene acts late in the splicing of yeast pre-messenger RNA, mediating the release of the spliced mRNA from the spliceosome	297	58
1478	ABB06166	Homo sapiens	AJIN Human liver with hepatitis related protein SEQ ID NO:12.	2688	100
1478	ABB06165	Homo sapiens	AJIN Human liver with hepatitis related protein SEQ ID NO:11.	1653	100
1478	gi9886891	Mus musculus	zinc finger protein 276 C2H2 type	462	35
1479	AAU76799	Homo sapiens	BLOW- Human oxysterol-binding protein 38_72.	1535	99
1479	gi6318855	Homo sapiens	A-kinase anchoring protein 220	379	22
1479	gi7671392	Homo sapiens	bA215B13.1 (A kinase (PRKA) anchor protein 11)	379	22
1480	gi13016757	Homo sapiens	dJ553F4.3 (novel zinc-finger protein)	2638	99
1480	AAB95176	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17237.	1130	100
1480	AAU16091	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1044.	848	80
1481	AAG00431	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 4512.	230	91
1481	gi20429112	Paracoccus zeaxanthinifaciens	mevalonate diphosphate decarboxylase	70	51
1481	gi17555680 [ref]NP_499688.1	Caenorhabditis elegans	Y37D8A.22.p	66	36
1482	AAM25955	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1470.	400	97
1482	gi14317890	Gallus gallus	spindlin	278	65
1482	gi17426439	Homo sapiens	bA445O16.1 (DXF34)	269	96
1483	AAR94963	Homo sapiens	INRM Human survival motor neuron SMN protein.	1580	90
1483	AAR96991	Homo sapiens	INRM Survival motor neuron protein, derived from clone T-BCD541.	1580	90
1483	gi624186	Homo sapiens	survival motor neuron	1580	90
1484	AAG02443	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6524.	304	100
1484	AAAY50923	Homo sapiens	ALPH- Human fetal brain cDNA clone vc10_1 derived protein.	68	48
1484	AAAY09001	Homo sapiens	USSH FMF associated protein pyrin MFEV.	66	40
1485	gi16755594	Bos taurus	mucin	110	30
1485	gi18495795	Bos taurus	MUC1 protein	110	30
1485	AAM41866	Homo sapiens	HYSE- Human polypeptide SEQ ID	108	30

286

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			NO 6797.		
1486	AAW19919	Homo sapiens	REGC Human Ksr-1 (kinase suppressor of Ras).	294	81
1486	gi1171250	Mus musculus	protein kinase related to Raf protein kinases; Method: conceptual translation supplied by author	277	78
1486	AAM25335	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:850.	221	76
1487	AAM69345	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29651.	3728	100
1487	AAM56960	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 29065.	3728	100
1487	AAM42395	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 128.	599	87
1488	AAM49117	Homo sapiens	CHUG- Human Sp1 family transcription factor h285.	2174	100
1488	gi9454416	Mus musculus	zinc finger protein Sp5	2122	97
1488	gi17978563	Xenopus laevis	Sp1-like zinc-finger protein XSPR-1	1306	65
1489	gi4323152	Mus musculus	Ets-protein Spi-C	876	67
1489	gi8745414	Aulonocara hansbaenschi	Spi-C transcription factor	467	49
1489	gi11245502	Raja eglanteria	SpiC	437	42
1490	AAG01095	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5176.	261	70
1490	gi7644350	Homo sapiens	golgi matrix protein GM130	192	69
1490	gi8099669	Homo sapiens	golgin-like protein	178	66
1491	gi 20903495 [ref]XP_128165.1]	Mus musculus	RIKEN cDNA 1700030F18	298	57
1491	gi 13541067 [ref]NP_110755.1]	Thermoplasma volcanium	Predicted DNA-binding protein	70	27
1492	AAM94393	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 3051.	367	95
1492	gi17861938	Drosophila melanogaster	HL03404p	238	43
1492	gi17862896	Drosophila melanogaster	SD02338p	238	43
1493	gi19527867	Drosophila melanogaster	AT11276p	257	26
1493	AAB94796	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15921.	252	26
1493	AAM39450	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2595.	252	26
1494	AAB93521	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12861.	581	41
1494	AAM90443	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:18036.	237	97
1494	AAM70186	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 30492.	166	46
1495	AAB92790	Homo sapiens	HELI- Human protein sequence SEQ	1709	68

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			ID NO:11286.		
1495	AAM87063	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:14656.	174	72
1495	AAM86989	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:14582.	162	84
1496	gi20810288	Homo sapiens	Similar to DnaJ (Hsp40) homolog, subfamily B, member 8	1236	100
1496	AAU17599	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 1164.	1232	99
1496	AAM95431	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4089.	1232	99
1497	gi8777659	Homo sapiens	transcription factor HOXD11	1576	89
1497	gi871428	Mus musculus	Hox-4.6	1549	86
1497	gi397509	Mus musculus	HOXD-11	1485	86
1498	gi7839650	Cricetulus griseus	plectin	127	20
1498	gi1296662	Homo sapiens	plectin	125	21
1498	gi1561642	Rattus norvegicus	plectin	119	20
1499	AAM79656	Homo sapiens	HYSE- Human protein SEQ ID NO 3302.	857	81
1499	AAM78672	Homo sapiens	HYSE- Human protein SEQ ID NO 1334.	857	81
1499	ABB12329	Homo sapiens	HYSE- Human Fas-ligand associated factor 3 homologue, SEQ ID NO:2699.	857	81
1500	gi12324505	Arabidopsis thaliana	AIG1-like protein; 41133-42535	188	27
1500	gi12324510	Arabidopsis thaliana	AIG1-like protein; 11637-17773	183	28
1500	gi4097585	Nicotiana tabacum	NTGP4	160	26
1501	gi1932775	Mus musculus	paired-type homeobox gene	276	35
1501	gi19354403	Mus musculus	retina and anterior neural fold homeobox	276	35
1501	gi2240024	Mus musculus	retinal homeobox protein	271	42
1502	AAU80387	Homo sapiens	CORI- Human lung tumour protein from clone L587S.	1320	100
1502	gi16041678	Homo sapiens	Similar to RIKEN cDNA 2810433K01 gene	1320	100
1502	AAU80388	Homo sapiens	CORI- Human lung tumour protein from recombinant L587S.	1315	100
1503	AAE01527	Homo sapiens	HUMA- Human gene 15 encoded secreted protein fragment, SEQ ID NO:184.	3358	100
1503	AAU21842	Homo sapiens	INCY- Human signal peptide-containing protein (SIGP) (clone ID 1273453).	1416	100
1503	AAE01494	Homo sapiens	HUMA- Human gene 15 encoded secreted protein HBKED12, SEQ ID NO:150.	879	99
1504	AAB95183	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17252.	1492	99
1504	AAG01491	Homo sapiens	GEST Human secreted protein, SEQ	752	99

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			ID NO: 5572.		
1504	gi949996	Rattus sp.	zinc finger protein RIZ	104	25
1505	gi5532389	Lytechinus variegatus	microtubule-associated protein	91	23
1505	gi19263503	Homo sapiens	Similar to RIKEN cDNA 1700019F09 gene	84	23
1505	gi3132790	Crithidia fasciculata	CACK protein	79	26
1506	AAB95409	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17793.	1040	100
1506	gi16716330	Canis familiaris	treacle	84	32
1506	gi20151453	Drosophila melanogaster	GH26759p	77	26
1507	gi11190493	Homo sapiens	bA345L23.2 (novel protein with BTB/POZ (broad complex Tramtrack bric-a-brac/Pox virus and zinc finger) domain)	3317	100
1507	gi6094684	Homo sapiens	similar to Kelch proteins; similar to BAA77027 (PID:g4650844)	973	34
1507	AAU28187	Homo sapiens	HYSE- Novel human secretory protein, Seq ID No 356.	948	35
1508	gi1478188	Gallus gallus	CHoxE	593	75
1508	gi6594621	Danio rerio	homeobox protein	315	57
1508	gi7689275	Xenopus laevis	homeodomain protein dbx	311	47
1509	AAG02366	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6447.	391	98
1509	gi16418290	Homo sapiens	RPGR-interacting protein 1	146	42
1509	gi9966409	Bos taurus	RPGR-interacting protein-1	107	42
1510	gi9884738	Homo sapiens	dJ983H21.2 (A novel protein similar to AP-2 beta transcription factor)	1151	100
1510	gi496639	Mus musculus	transcription factor Ap-2 beta	1036	77
1510	gi1495417	Homo sapiens	AP-2 beta	1033	76
1511	AAB95825	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18834.	1429	100
1511	gi2605967	Equine herpesvirus 4	24	156	25
1511	gi21324755	Corynebacterium glutamicum ATCC 13032	Translation initiation factor 2 (GTPase)	146	26
1512	AAB75305	Homo sapiens	ROSE/ Human secreted protein sequence encoded by gene 13 SEQ ID NO:124.	2785	98
1512	gi19528255	Drosophila melanogaster	GH13383p	2089	56
1512	gi987227	Caenorhabditis elegans	Nucampholin	1822	46
1513	ABB05592	Homo sapiens	UYNA- Mitotic centromere-associated kinesin NYD-KIF2 related fusion protein.	3802	99
1513	ABB05589	Homo sapiens	UYNA- Human testis-specific protein 1 NYD-TSP1 related fusion protein.	3802	99
1513	AAU10784	Homo sapiens	UYNA- Human ubiquitin-like fusing protein (UFLP).	3802	99
1514	gi16755530	Homo sapiens	androgen receptor-associated coregulator 267-a	12441	99

Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1514	gi16751269	Homo sapiens	androgen receptor associated coregulator 267-b	12441	99
1514	gi15213542	Homo sapiens	NSD1	11819	97
1515	AAB94881	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16098.	1404	99
1515	AAB92639	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10961.	94	30
1515	AAG89192	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 312.	94	30
1516	gi7159336	Entamoeba histolytica	diaphanous protein	72	29
1516	gi3421373	Mus musculus	28 kDa cis-Golgi SNARE	71	25
1516	gi14250239	Mus musculus	golgi SNAP receptor complex member 1	71	25
1517	gi15929485	Homo sapiens	Similar to RIKEN cDNA 4930429A08 gene	2144	100
1517	AAB94367	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14901.	840	99
1517	AAM82536	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:10129.	433	85
1518	AAB95217	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17333.	89	24
1518	gi20515041	Thermoanaerobacter tengcongensis	nucleoside-diphosphate-sugar pyrophosphorylase	86	27
1518	gi854353	Oryctolagus cuniculus	neurofilament protein M	85	27
1519	AAB94133	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14393.	3217	99
1519	AAG81359	Homo sapiens	ZYMO Human AFP protein sequence SEQ ID NO:236.	1874	98
1519	gi17944262	Drosophila melanogaster	LD30968p	901	30
1520	gi10440377	Homo sapiens	FLJ00024 protein	695	71
1520	AAB64372	Homo sapiens	INCY- Amino acid sequence of human intracellular signalling molecule INTRA4.	83	33
1520	AAG74084	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4848.	83	33
1521	AAM77456	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 37762.	462	100
1521	AAM64678	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 36783.	462	100
1521	AAM68208	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 28514.	448	100
1522	AAM25562	Homo sapiens	HYSE- Human protein sequence SEQ ID NO:1077.	1299	96
1522	AAV73335	Homo sapiens	INCY- HTRM clone 1850120 protein sequence.	1294	95
1522	gi1817526	Anthocidaris crassispina	intermediate chain 1	142	30

290
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1523	gi1310709	Rattus norvegicus	sca1	667	29
1523	gi2808423	Homo sapiens	ataxin-1 (spinocerebellar ataxia type 1 protein)	659	29
1523	AAY33494	Homo sapiens	BURN- Human SCA1 protein.	658	29
1524	AAG04066	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8147.	414	97
1524	gi16185010	Drosophila melanogaster	LD32609p	72	37
1525	gi19919744	Homo sapiens	hepatocellular carcinoma-associated protein HCA3	1362	100
1525	gi12659148	Mus musculus	mage-e2	656	88
1525	gi19919740	Homo sapiens	hepatocellular carcinoma-associated protein HCA1	520	38
1526	gi4336205	Zea mays	cytochrome b5 reductase	271	31
1526	gi14536592	Physcomitrella patens	PP001069030R	270	31
1526	gi14536588	Physcomitrella patens	25_ppprot1_046_e01	263	36
1527	AAU82007	Homo sapiens	INCY- Human secreted protein SECP33.	1213	47
1527	AAM39715	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2860.	1213	47
1527	AAM41501	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6432.	1193	48
1528	AAU82007	Homo sapiens	INCY- Human secreted protein SECP33.	1207	48
1528	AAM39715	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2860.	1207	48
1528	AAM41501	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6432.	1187	48
1529	gi3253120	Homo sapiens	R31449_3	2305	99
1529	gi7673675	Drosophila melanogaster	cactin	1038	42
1529	gi21464388	Drosophila melanogaster	RE14858p	839	45
1530	AAM66321	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 26627.	467	76
1530	AAM53933	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 26038.	467	76
1530	AAM75161	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 35467.	314	78
1531	gi8346962	Mus musculus	1A13 protein	1990	97
1531	AAB93285	Homo sapiens	HELI- Human protein sequence SEQ ID NO:12336.	1182	59
1531	AAM79275	Homo sapiens	HYSE- Human protein SEQ ID NO 1937.	1182	59
1532	gi2252814	Mus musculus	FOG	3629	69
1532	gi7595837	Xenopus laevis	Friend of GATA	1847	43
1532	gi4927696	Mus musculus	zinc-finger factor FOG2	949	31
1533	gi4079709	Rattus norvegicus	reggie1-1	2106	99

291
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1533	gi4079711	Rattus norvegicus	reggie1-2	2063	96
1533	gi13277550	Homo sapiens	Similar to flotillin 2	1862	99
1534	gi16118243	Homo sapiens	ARAP1	6340	100
1534	ABB07500	Homo sapiens	INCY- Human GTP-binding protein (GTPB) (ID: 1299273CD1).	6287	99
1534	gi15625574	Homo sapiens	centaurin delta2	5951	100
1535	gi16118243	Homo sapiens	ARAP1	6255	99
1535	ABB07500	Homo sapiens	INCY- Human GTP-binding protein (GTPB) (ID: 1299273CD1).	6202	98
1535	gi15625574	Homo sapiens	centaurin delta2	5866	99
1536	AAU17145	Homo sapiens	HUMA- Novel signal transduction pathway protein, Seq ID 710.	630	100
1536	AAM48374	Homo sapiens	SHAN- Human PP905 protein.	157	41
1536	gi10834732	Homo sapiens	PP905	157	41
1537	gi20988071	Mus musculus	Similar to RIKEN cDNA 2600011E07 gene	156	47
1537	gi6958206	Pneumocystis carinii f. sp. muris	kexin-like protease KEX1	137	30
1537	gi15917538	Homo sapiens	NG36/G9a	131	26
1538	gi17946424	Drosophila melanogaster	RE67445p	241	42
1538	gi21464585	Arabidopsis thaliana	AT3g02910/F13E7_14	181	42
1538	gi552255	Lytechinus pictus	troponin C	146	51
1539	ABB11447	Homo sapiens	HYSE- Human secreted protein homologue, SEQ ID NO:1817.	1764	100
1539	ABB89981	Homo sapiens	HUMA- Human polypeptide SEQ ID NO 2357.	1634	100
1539	AAAY41528	Homo sapiens	HUMA- Fragment of human secreted protein encoded by gene 77.	1172	99
1540	AAB95462	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17944.	1488	100
1540	AAB94494	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15186.	1246	99
1540	AAG64392	Homo sapiens	BIOD- Human reducing agent and tunicamycin-responsive protein 40.	1246	99
1541	gi14089673	Mycoplasma pulmonis	50S RIBOSOMAL PROTEIN L35	229	74
1541	gi150150	Mycoplasma fermentans	ribosomal protein L35	213	69
1541	gi21204792	Staphylococcus aureus subsp. aureus MW2	50S ribosomal protein L35	175	58
1542	gi13543686	Homo sapiens	Similar to RIKEN cDNA 4931428F02 gene	750	97
1542	AAAY07081	Homo sapiens	LUDW- Renal cancer associated antigen precursor sequence.	261	38
1542	AAE13761	Homo sapiens	CORI- Human lung tumour-specific protein LT86-14.	226	35
1543	gi19683951	Homo sapiens	Similar to DNA segment, Chr 6, Miriam Meisler 5, expressed	2234	99
1543	gi4008004	Mus musculus	D6MM5E protein	1651	74

292
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1543	gi4007988	Mus musculus	D6MM5e protein	1651	74
1544	gi11493263	Homo sapiens	bA196N14.1 (novel protein)	749	100
1544	gi1652299	Synechocystis sp. PCC 6803	DNA ligase	77	32
1544	gi6492366	Mus musculus	CLNK	72	36
1545	gi13938461	Homo sapiens	Similar to RIKEN cDNA 4933428D01 gene	938	100
1545	AAM79504	Homo sapiens	HYSE- Human protein SEQ ID NO 3150.	90	28
1545	AAM78520	Homo sapiens	HYSE- Human protein SEQ ID NO 1182.	90	28
1546	AAB94901	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16290.	813	58
1546	gi13122230	Leishmania major	CDC27/NUC2-related protein	101	28
1546	gi468012	Bos taurus	PKR inhibitor P58	100	26
1547	gi18568414	Homo sapiens	BRG1-binding protein ELD/OSA1	3291	83
1547	gi11527189	Homo sapiens	p250R	2021	84
1547	gi8489817	Homo sapiens	SWI-SNF complex protein p270	1689	47
1548	ABB06069	Homo sapiens	COMP- Human NS protein sequence SEQ ID NO:161.	1342	95
1548	AAG89296	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 416.	1342	95
1548	AAG89265	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 385.	1330	95
1549	gi9837427	Lytechinus variegatus	embryonic blastocoelar extracellular matrix protein precursor	259	29
1549	gi15291905	Drosophila melanogaster	LD31354p	152	29
1549	gi1002808	Lytechinus variegatus	extracellular matrix protein	139	25
1550	gi7288419	Treponema pallidum	tpr protein K	68	48
1550	gi 11095200 gb AAG29772.1 AF227220_1	Treponema pallidum	TprK	66	45
1552	gi3510757	Homo sapiens	glutathione transferase zeta 1	848	80
1552	gi7417477	Homo sapiens	GTZ1	848	80
1552	gi12655191	Homo sapiens	glutathione transferase zeta 1 (maleylacetoacetate isomerase)	848	80
1553	AAB94701	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15686.	858	100
1553	AAM79264	Homo sapiens	HYSE- Human protein SEQ ID NO 1926.	171	22
1553	gi7839650	Cricetulus griseus	plectin	171	22
1554	gi9971158	Homo sapiens	GTP-binding like protein 2	1286	99
1554	gi13561007	Homo sapiens	bA22I24.2.1 (GTP binding protein 2)	1286	99
1554	gi20306212	Homo sapiens	GTP binding protein 2	1285	99
1555	AAG93258	Homo sapiens	NISC- Human protein HP10582.	1348	100
1555	AAG74631	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:5395.	869	95
1555	gi17863052	Drosophila melanogaster	SD10403p	350	35

293
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1556	AAY92330	Homo sapiens	CURA- Human NIK1 protein.	1918	87
1556	gi507875	Homo sapiens	NIMA-like protein kinase 1	1918	87
1556	gi479171	Homo sapiens	protein kinase	1918	87
1557	AAY92330	Homo sapiens	CURA- Human NIK1 protein.	1943	89
1557	gi507875	Homo sapiens	NIMA-like protein kinase 1	1943	89
1557	gi479171	Homo sapiens	protein kinase	1943	89
1558	gi13469818	Mus musculus	bicaudal-C	2179	89
1558	gi7800180	Xenopus laevis	bicaudal-C	1697	73
1558	gi13592175	Leishmania major	ppg3	138	20
1559	AAB95608	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18306.	1976	99
1559	AAB94407	Homo sapiens	HELI- Human protein sequence SEQ ID NO:14989.	1847	98
1559	ABB04606	Homo sapiens	BODA- Human RNA recombining area 18 protein SEQ ID NO:2.	849	100
1560	gi15823651	Homo sapiens	ALS2CR11	1254	82
1560	AAM57442	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 29547.	809	88
1560	AAV83134_aa1	Homo sapiens	GEMY Polynucleotide clone du410_5 encoding a secreted protein.	80	25
1561	gi5869804	Drosophila melanogaster	cramped protein	287	25
1561	gi15291845	Drosophila melanogaster	LD29481p	287	25
1561	gi557822	Saccharomyces cerevisiae	mal5, sta1, len: 1367, CAI: 0.3, AMYH_YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	175	20
1562	AAB95707	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18550.	896	100
1562	gi15026974	Homo sapiens	obscurin	449	96
1562	AAU25456	Homo sapiens	INCY- Human mddt protein from clone LI:406860.20:2000MAY01.	149	36
1563	AAU11805	Homo sapiens	MDSP- Human GRF2 protein.	6402	100
1563	gi5882290	Homo sapiens	Ras guanine nucleotide exchange factor 2	6402	100
1563	AAM39825	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2970.	6397	99
1564	AAU11805	Homo sapiens	MDSP- Human GRF2 protein.	6412	100
1564	gi5882290	Homo sapiens	Ras guanine nucleotide exchange factor 2	6412	100
1564	AAM39825	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2970.	6407	99
1565	gi951270	Homo sapiens	IPP-2	503	79
1565	gi474388	Homo sapiens	inhibitor 2	503	79
1565	gi4704218	Homo sapiens	inhibitor 2 of protein phosphatase 1	503	79
1566	gi20152557	Mus musculus	dendritic cell and B cell derived chemokine	72	25
1566	gi5911365	Mus musculus	macrophage-derived chemokine	72	25
1566	gi5757694	Mus musculus	macrophage-derived chemokine MDC/ABCD-1	72	25
1567	gi13676773	Chlamydomonas reinhardtii	PF6 protein	127	31
1567	AAM79809	Homo sapiens	HYSE- Human protein SEQ ID NO	111	26

294
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			3455.		
1567	AAM78825	Homo sapiens	HYSE- Human protein SEQ ID NO 1487.	111	26
1568	AAB95872	Homo sapiens	HELI- Human protein sequence SEQ ID NO:18952.	1641	100
1568	gi9800317	rat cytomegalovirus Maastricht	pR102	141	32
1568	gi13562004	Nephila madagascariensis	major ampullate spidroin 2-like protein	138	32
1569	AAU07889	Homo sapiens	WHED Polypeptide sequence for human hspG34a.	963	92
1569	AAU07890	Homo sapiens	WHED Polypeptide sequence for human hspG34b.	575	72
1569	gi13603867	Homo sapiens	ferritin heavy polypeptide-like 17	575	72
1570	AAU11817	Homo sapiens	UYLE- Cancer and neurogenesis associated gene, variant 5R23V2.	1577	69
1570	AAU11816	Homo sapiens	UYLE- Cancer and neurogenesis associated gene, variant 5R-3V2.	1577	69
1570	AAU11814	Homo sapiens	UYLE- Cancer and neurogenesis associated gene, variant 5G-3V2.	1577	69
1571	gi18204103	Mus musculus	RIKEN cDNA 4930429H24 gene	531	38
1571	AAM38956	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2101.	496	31
1571	gi6644176	Homo sapiens	kelch-like protein KLHL3a	496	31
1572	AAO12605	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 26497.	941	100
1572	gi15145797	Sus scrofa	basic proline-rich protein	402	32
1572	gi5917666	Zea mays	extensin-like protein	375	29
1573	AAG75965	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6729.	456	96
1573	AAG01569	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5650.	456	96
1573	AAB43549	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:994.	456	96
1574	AAB36610	Homo sapiens	INCY- Human FLEXHT-32 protein sequence SEQ ID NO:32.	547	46
1574	AAM39124	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 2269.	547	46
1574	gi6841176	Homo sapiens	HSPC263	547	46
1575	AAG75449	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:6213.	1070	99
1575	gi2982673	Homo sapiens	p27	1033	100
1575	gi15928646	Homo sapiens	Sjogren's syndrome/scleroderma autoantigen 1	1033	100
1576	gi19548926	Homo sapiens	rhysin 2	2721	99
1576	AAM00777	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 140.	2659	99
1576	AAM00890	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 366.	1134	100
1577	ABB50256	Homo sapiens	INCY- Human transcription factor TRFX-107.	2728	99
1577	gi20987263	Homo sapiens	zinc finger protein 289, ID1 regulated	2636	97
1577	AAB95129	Homo sapiens	HELI- Human protein sequence SEQ ID NO:17133.	2625	96

295
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1578	gi9622219	Rattus norvegicus	beta-catenin binding protein	2550	95
1578	AAM76258	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 36564.	560	100
1578	AAM63443	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID NO: 35548.	560	100
1579	gi3868802	Mus musculus	c29	1704	73
1579	gi12642308	Ovis aries	type I keratin intermediate filament IRSa1	1671	72
1579	AAB34359	Homo sapiens	HUMA- Human secreted protein sequence encoded by gene 6 SEQ ID NO:120.	1228	69
1580	AAV83093	Homo sapiens	UYNF F-box protein FBP-25.	1002	100
1580	gi6164743	Homo sapiens	F-box protein Fbx20	1002	100
1580	AAV83072	Homo sapiens	UYNF F-box motif of FBP-25.	261	100
1581	gi2622381	Methanothermobacter thermotrophicus str. Delta H	conserved protein	107	37
1581	gi19888349	Methanopyrus kandleri AV19	Zn-dependent hydrolase	80	31
1581	gi5458479	Pyrococcus abyssi	HYDROXYACYLGLUTATHIONE HYDROLASE related (EC 3.1.2.6) (GLYOXALASE II)	79	26
1582	gi12655474	Homo sapiens	keratin associated protein 9.8	961	94
1582	gi12655470	Homo sapiens	keratin associated protein 9.4	933	93
1582	gi12655466	Homo sapiens	keratin associated protein 9.2	899	85
1583	gi6018193	Mus musculus	nucleotide-binding protein long form	1391	82
1583	AAV73353	Homo sapiens	INCY- HTRM clone 1870914 protein sequence.	1344	83
1583	gi19570100	Dictyostelium discoideum	Nucleotide-binding protein 1	978	60
1584	gi20071260	Mus musculus	RIKEN cDNA 2310038H17 gene	242	35
1584	gi18252514	Homo sapiens	hepatocellular carcinoma-associated antigen HCA557b	233	34
1584	gi18252512	Homo sapiens	hepatocellular carcinoma-associated antigen HCA557a	213	31
1585	AAM00916	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 392.	1060	40
1585	AAM00803	Homo sapiens	HYSE- Human bone marrow protein, SEQ ID NO: 166.	1060	40
1585	AAM79000	Homo sapiens	HYSE- Human protein SEQ ID NO 1662.	1060	40
1586	AAG03361	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7442.	204	97
1586	gi19481932	shrimp white spot syndrome virus	WSSV340	80	41
1586	gi17016682	shrimp white spot syndrome virus	wsv285	80	41
1587	gi3213225	Homo sapiens	T-box-containing transcriptional activator	952	84

296
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
1587	gi12620817	Mus musculus	T-box 1	799	57
1587	gi13173432	Mus musculus	T-box 1 transcription factor	791	56
1588	ABB55706	Homo sapiens	FECH/ Human polypeptide SEQ ID NO 18.	102	24
1588	AAU38997	Homo sapiens	GEMY Human secreted protein ya11_1.	102	24
1588	AAU17227	Homo sapiens	GEMY Human secreted protein (clone ya1-1).	102	24
1589	gi14577933	Mus musculus	ribonuclease/angiogenesis inhibitor	258	29
1589	AAB14966	Homo sapiens	WISC Human ribonuclease inhibitor.	247	29
1589	gi190848	Homo sapiens	ribonuclease/angiogenin inhibitor	247	29
1590	AAB43434	Homo sapiens	HUMA- Human cancer associated protein sequence SEQ ID NO:879.	378	74
1590	AAU07039	Homo sapiens	LUDW- Breast cancer associated antigen precursor sequence.	378	74
1590	gi190234	Homo sapiens	acidic ribosomal phosphoprotein (P1)	378	74
1591	gi3171877	Homo sapiens	dJ127D3.2 (Flavin-containing Monooxygenase family protein)	2862	97
1591	AAW49700	Homo sapiens	INRM Human flavin-containing monooxygenase isoform x.	2794	96
1591	gi12006730	Rattus norvegicus	flavin-containing monooxygenase FMO3	2138	72
1592	gi3171877	Homo sapiens	dJ127D3.2 (Flavin-containing Monooxygenase family protein)	2862	97
1592	AAW49700	Homo sapiens	INRM Human flavin-containing monooxygenase isoform x.	2794	96
1592	gi12006730	Rattus norvegicus	flavin-containing monooxygenase FMO3	2138	72
1593	AAG74122	Homo sapiens	HUMA- Human colon cancer antigen protein SEQ ID NO:4886.	375	83
1593	gi6633816	Arabidopsis thaliana	F1N19.14	140	27
1593	gi9837385	Takifugu rubripes	retinitis pigmentosa GTPase regulator-like protein	130	21
1594	AAB92569	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10777.	1911	100
1594	gi531218	Rattus norvegicus	amino acid feature: homeodomain, bp 647 .. 826; amino acid feature: LIM1, bp 155 .. 307; amino acid feature: LIM2, bp 341 .. 493	1828	96
1594	gi1037166	Danio rerio	zfIsl-2	1683	88
1595	AAB94572	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15358.	736	100
1596	AAG02673	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 6754.	359	100
1596	gi9916	Plasmodium falciparum	liver stage antigen	300	20
1596	AAM82393	Homo sapiens	HUMA- Human immune/haematopoietic antigen SEQ ID NO:9986.	223	95
1597	AAM69464	Homo sapiens	MOLE- Human bone marrow expressed probe encoded protein SEQ ID NO: 29770.	456	98
1597	AAM57075	Homo sapiens	MOLE- Human brain expressed single exon probe encoded protein SEQ ID	456	98

297
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
			NO: 29180.		
1597	gi18676566	Homo sapiens	FLJ00180 protein	272	31
1598	AAG03347	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 7428.	219	97
1598	AAB92646	Homo sapiens	HELI- Human protein sequence SEQ ID NO:10979.	117	39
1598	AAU74388	Homo sapiens	CORI- Breast tumour-specific protein B305D.	112	38
1599	AAM42062	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6993.	346	67
1599	AAB08765	Homo sapiens	INCY- A human leukocyte and blood related protein (LBAP).	278	60
1599	AAM40276	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 3421.	274	75
1600	gi11385354	Homo sapiens	BAF180	8308	100
1600	gi12083875	Homo sapiens	polybromo-1	8206	96
1600	AAG78407	Homo sapiens	MERE Amino acid sequence of hupolybromo1.	8135	96
1601	AAB94448	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15085.	2590	99
1601	gi10440369	Homo sapiens	FLJ00019 protein	2370	99
1601	gi2344894	Arabidopsis thaliana	F-box protein family, AtFBX5	137	24
1602	gi19918958	Homo sapiens	AMAP-1	4919	99
1602	ABB05712	Homo sapiens	GEHU- Human testis derived protein clone tes3_15n14.	3720	100
1602	AAM95134	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 3792.	1005	98
1603	AAM95435	Homo sapiens	HUMA- Human reproductive system related antigen SEQ ID NO: 4093.	703	99
1603	AAG01839	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 5920.	444	95
1603	gi20380952	Mus musculus	Similar to RIKEN cDNA 1700001C19 gene	413	68
1604	gi1199604	Homo sapiens	zinc finger protein C2H2-25	2055	100
1604	AAM78976	Homo sapiens	HYSE- Human protein SEQ ID NO 1638.	1652	55
1604	AAM79960	Homo sapiens	HYSE- Human protein SEQ ID NO 3606.	1630	55
1605	gi22380	Zea mays	CAAT-box DNA binding protein subunit B (NF-YB)	72	32
1605	gi14970736 emb CAC4 4453.1	Danio rerio	insulin-like growth factor binding protein 1	68	33
1606	AAM79754	Homo sapiens	HYSE- Human protein SEQ ID NO 3400.	173	32
1606	AAM78770	Homo sapiens	HYSE- Human protein SEQ ID NO 1432.	168	33
1606	gi17226682	Anopheles gambiae	Trex	106	27
1607	gi4929701	Homo sapiens	CGI-116 protein	755	82
1607	gi7677064	Homo sapiens	protein x 0009	755	82
1607	gi15011990	Homo sapiens	Similar to CGI-116 protein	750	82
1608	gi11640582	Homo sapiens	MSTP037	687	100
1608	gi5305335	Mycobacterium	proline-rich mucin homolog	137	39

298
Table 2B

SEQ ID	Hit ID	Species	Description	S score	Percent identity
		tuberculosis			
1608	gi7363175	Schizosaccharomyces pombe	actin binding protein with SH3 domains	130	33
1609	gi2138290	Homo sapiens	see GenBank Accession Number U01184 for cDNA; similar to Drosophila melanogaster fliI in GenBank Accession Number U01182 and Caenorhabditis elegans fliI homolog in GenBank Accession Number U01183	6615	98
1609	gi19263700	Homo sapiens	flightless I homolog (Drosophila)	6615	98
1609	gi440177	Homo sapiens	flightless-I homolog	6610	98
1610	gi4808631	Homo sapiens	transcription factor-like protein MRGX	532	94
1610	gi8895210	Homo sapiens	MSL3-2 protein	532	94
1610	gi5931553	Mus musculus	Sid393p	482	85
1611	gi12697482	Homo sapiens	dJ583P15.7.2 (novel zinc finger protein similar to rat RIN ZF)	1179	100
1611	gi9843768	Homo sapiens	dJ583P15.7.1 (novel zinc finger protein similar to rat RIN ZF)	924	100
1611	AAU16398	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 1351.	810	99
1612	gi18043255	Mus musculus	Similar to RIKEN cDNA 4733401D09 gene	1590	53
1612	gi14714943	Homo sapiens	Similar to outer dense fiber of sperm tails 2	698	28
1612	gi17388906	Rattus norvegicus	cenexin 2	604	27
1613	gi18043255	Mus musculus	Similar to RIKEN cDNA 4733401D09 gene	1926	64
1613	gi14714943	Homo sapiens	Similar to outer dense fiber of sperm tails 2	604	27
1613	gi17388906	Rattus norvegicus	cenexin 2	508	25
1614	AAM93657	Homo sapiens	HELI- Human polypeptide, SEQ ID NO: 3528.	2580	99
1614	gi10177621	Arabidopsis thaliana	phytoene dehydrogenase-like	1103	49
1614	gi17979255	Arabidopsis thaliana	AT5g49550/K6M13_10	1099	49
1615	AAU15903	Homo sapiens	HUMA- Human novel secreted protein, Seq ID 856.	2328	86
1615	AAB94636	Homo sapiens	HELI- Human protein sequence SEQ ID NO:15515.	2108	100
1615	gi20197056	Arabidopsis thaliana	expressed protein	332	28
1616	AAG04014	Homo sapiens	GEST Human secreted protein, SEQ ID NO: 8095.	281	92
1617	gi21309945	Mus musculus	MRS3/4	378	85
1617	AAM41505	Homo sapiens	HYSE- Human polypeptide SEQ ID NO 6436.	286	88
1617	gi16755528	Mus musculus	mitochondrial carrier-like protein	159	54
1619	AAB94885	Homo sapiens	HELI- Human protein sequence SEQ ID NO:16106.	1454	99
1619	gi19483920	Mus musculus	RIKEN cDNA 2310008M20 gene	1224	82
1619	AAG01968	Homo sapiens	GEST Human secreted protein, SEQ	644	94